GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 22, 2005, 10:02:08 ; Search time 178 Seconds
(without alignments)
1668.573 Million cell updates/sec

Title: US-10-629-951-2
Perfect score: 3055
Sequence: 1 MGTTGLESLSLGDRGAAPTV.....HKLNKYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3055	100.0	580	1	MBD4_HUMAN	O95243 homo sapien
2	1821.5	59.6	554	1	MBD4_MOUSE	Q9z2d7 mus musculu
3	880.5	28.8	416	2	Q9I9F1	Q9i9f1 gallus gall
4	344	11.3	419	2	Q9SFC1	Q9sfc1 arabidopsis
5	338	11.1	407	2	Q84WT3	Q84wt3 arabidopsis
6	238.5	7.8	524	2	Q7T2T7	Q7t2t7 brachydanio
7	220.5	7.2	467	2	Q9YGC6	Q9ygc6 xenopus lae
8	213	7.0	682	2	Q7SCQ2	Q7scq2 neurospora
9	208.5	6.8	486	1	MEC2_HUMAN	P51608 homo sapien
10	208.5	6.8	498	2	Q6QHH9	Q6qhh9 homo sapien
11	208.5	6.8	516	2	Q7Z384	Q7z384 homo sapien
12	207.5	6.8	486	1	MEC2_MACFA	Q95lg8 macaca fasc
13	202.5	6.6	344	2	O42403	O42403 gallus gall
14	201	6.6	492	1	MEC2_RAT	Q00566 rattus norv
15	198.5	6.5	484	1	MEC2_MOUSE	Q9z2d6 mus musculu
16	156	5.1	2473	2	Q7RN41	Q7rn41 plasmodium
17	149.5	4.9	2212	2	Q8I1T6	Q8i1t6 plasmodium
18	148.5	4.9	352	2	Q8LB76	Q8lb76 arabidopsis
19	148.5	4.9	1679	2	Q7RCP2	Q7rcp2 plasmodium
20	146	4.8	702	2	Q9SF22	Q9sf22 arabidopsis
21	146	4.8	1643	2	Q8HZQ1	Q8hzq1 pan troglod
22	145	4.7	267	2	Q8AYT1	Q8ayt1 xenopus lae
23	145	4.7	1301	1	SAC3_YEAST	P46674 saccharomyc
24	145	4.7	3418	1	BRC2_HUMAN	P51587 homo sapien
25	144	4.7	283	2	Q8AYP2	Q8ayp2 xenopus lae
26	143.5	4.7	782	2	Q25875	Q25875 plasmodium
27	141.5	4.6	782	2	Q25730	Q25730 plasmodium
28	141.5	4.6	782	2	Q26007	Q26007 plasmodium
29	140.5	4.6	782	2	Q9U414	Q9u414 plasmodium
30	140.5	4.6	782	2	Q9U430	Q9u430 plasmodium
31	140.5	4.6	881	2	Q7RK82	Q7rk82 plasmodium
32	140.5	4.6	1030	2	Q924D9	Q924d9 meriones sh
33	138.5	4.5	500	1	GAR2_SCHPO	P41891 schizosacch
34	138.5	4.5	782	2	Q9U429	Q9u429 plasmodium
35	138.5	4.5	1047	2	Q7RPJ9	Q7rpj9 plasmodium
36	137.5	4.5	326	2	Q869D5	Q869d5 branchiosto
37	137.5	4.5	412	2	Q6MFI1	Q6mfi1 neurospora
38	137.5	4.5	782	2	Q26104	Q26104 plasmodium
39	137.5	4.5	2209	2	Q9U0G6	Q9u0g6 plasmodium
40	137	4.5	400	2	Q8LK06	Q8lk06 zea mays (m
41	137	4.5	560	1	SET4_YEAST	P42948 saccharomyc
42	137	4.5	707	2	Q7RKH5	Q7rkh5 plasmodium
43	137	4.5	1713	2	Q7TT50	Q7tt50 mus musculu
44	137	4.5	1813	2	Q75XE3	Q75xe3 helicobacte
45	136.5	4.5	598	1	CYL1_HUMAN	P35663 homo sapien

ALIGNMENTS

RESULT 1
MBD4_HUMAN
ID MBD4_HUMAN STANDARD; PRT; 580 AA.
AC O95243; Q7Z4T3; Q96F09;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Methyl-CpG binding protein 4 (EC 3.2.2.-) (Methyl-CpG binding domain
DE protein 4) (Methyl-CpG binding endonuclease 1) (Mismatch-specific DNA
DE N-glycosylase).
GN Name=MBD4; Synonyms=MED1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
RT CpG binding proteins.";
RL Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S.H., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
RT mbd1, mbd2, mbd3, and mbd4 genes.";
RL Mamm. Genome 10:906-912(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INTERACTION WITH MLH1.
RC TISSUE=Fetal brain;
RX MEDLINE=99199294; PubMed=10097147; DOI=10.1073/pnas.96.7.3969;
RA Bellacosa A., Cicchillitti L., Schepis F., Riccio A., Yeung A.T.,
RA Matsumoto Y., Golemis E.A., Genuardi M., Neri G.;
RT "MED1, a novel human methyl-CpG-binding endonuclease, interacts with
RT DNA mismatch repair protein MLH1.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:3969-3974(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Lung;
RA Guo J.H., Chen L., Yu L.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS SER-273; PRO-342; LYS-346
RP AND HIS-568.
RA Rieder M.J., Braun A.C., Montoya M.A., Chung M.-W., Nguyen C.P.,
RA Nguyen D.A., Livingston R.J., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RT "NIEHS-SNPs, environmental genome project, NIEHS ES15478, Department
RT of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 2).

```
RA   Ebert L., Schick M., Neubert P., Schatten R., Henze S., Korn B.;
RT   "Cloning of human full open reading frames in Gateway(TM) system entry
RT   vector (pDONR201).";
RL   Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [8]
RP   FUNCTION.
RX   PubMed=10930409; DOI=10.1074/jbc.M004535200;
RA   Petronzelli F., Riccio A., Markham G.D., Seeholzer S.H., Stoerker J.,
RA   Genuardi M., Yeung A.T., Matsumoto Y., Bellacosa A.;
RT   "Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a
RT   mismatch-specific DNA N-glycosylase.";
RL   J. Biol. Chem. 275:32422-32429(2000).
RN   [9]
RP   INTERACTION WITH FADD.
RX   PubMed=12702765; DOI=10.1073/pnas.0431215100;
RA   Screaton R.A., Kiessling S., Sansom O.J., Millar C.B., Maddison K.,
RA   Bird A., Clarke A.R., Frisch S.M.;
RT   "Fas-associated death domain protein interacts with methyl-CpG binding
RT   domain protein 4: a potential link between genome surveillance and
RT   apoptosis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:5211-5216(2003).
CC   -!- FUNCTION: Mismatch-specific DNA N-glycosylase involved in DNA
CC       repair. Has thymine glycosylase activity and is specific for G:T
CC       mismatches within methylated and unmethylated CpG sites. Can also
CC       remove uracil or 5-fluorouracil in G:U mismatches. Has no lyase
CC       activity. Was first identified as methyl-CpG-binding protein.
CC   -!- SUBUNIT: Interacts with MLH1.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=O95243-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O95243-2; Sequence=VSP_010816;
CC         Note=No experimental confirmation available;
CC       Name=3;
CC         IsoId=O95243-3; Sequence=VSP_010817, VSP_010818;
CC         Note=No experimental confirmation available;
CC   -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072250; AAC68879.1; -.
DR   EMBL; AF120999; AAD50374.1; -.
DR   EMBL; AF120997; AAD50374.1; JOINED.
DR   EMBL; AF120998; AAD50374.1; JOINED.
DR   EMBL; AF114784; AAD22195.1; -.
DR   EMBL; AF532602; AAP97338.1; -.
DR   EMBL; AF494057; AAM00008.1; -.
DR   EMBL; CR450305; CAG29301.1; -.
DR   EMBL; BC011752; AAH11752.1; -.
DR   HSSP; Q9Z2D7; 1NGN.
DR   IntAct; O95243; -.
DR   Genew; HGNC:6919; MBD4.
DR   H-InvDB; HIX0003669; -.
DR   Reactome; O95243; -.
DR   MIM; 603574; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0004520; F:endodeoxyribonuclease activity; TAS.
DR   GO; GO:0003696; F:satellite DNA binding; TAS.
DR   GO; GO:0006281; P:DNA repair; TAS.
DR   InterPro; IPR003265; Endo_3c.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF00730; HhH-GPD; 1.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
DR   PROSITE; PS50982; MBD; 1.
KW   Alternative splicing; DNA repair; DNA-binding; Hydrolase;
KW   Nuclear protein; Polymorphism.
FT   DOMAIN       76    148       MBD.
FT   DOMAIN      461    524       HhH-GPD.
FT   ACT_SITE    560    560       By similarity.
FT   VARSPLIC    395    400       Missing (in isoform 2).
FT                                /FTId=VSP_010816.
FT   VARSPLIC    539    540       KY -> AP (in isoform 3).
FT                                /FTId=VSP_010817.
FT   VARSPLIC    541    580       Missing (in isoform 3).
FT                                /FTId=VSP_010818.
FT   VARIANT     273    273       A -> S (in dbSNP:10342).
FT                                /FTId=VAR_019357.
FT   VARIANT     273    273       A -> T (in dbSNP:10342).
FT                                /FTId=VAR_019514.
FT   VARIANT     342    342       S -> P (in dbSNP:2307289).
FT                                /FTId=VAR_019358.
FT   VARIANT     346    346       E -> K (in dbSNP:140693).
FT                                /FTId=VAR_019359.
FT   VARIANT     358    358       I -> T (in dbSNP:2307298).
FT                                /FTId=VAR_019515.
FT   VARIANT     568    568       D -> H (in dbSNP:2307293).
FT                                /FTId=VAR_019360.
SQ   SEQUENCE   580 AA;  66050 MW;  BF16FB21A34B8E5F CRC64;

  Query Match             100.0%;  Score 3055;  DB 1;  Length 580;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-178;
  Matches  580;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
```

```
Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              ||||||||||||||||||||||||||||||||||||||||
Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
```

```
RESULT 2
MBD4_MOUSE
ID   MBD4_MOUSE     STANDARD;      PRT;   554 AA.
AC   Q9Z2D7; Q792D2; Q8R3R3;
DT   25-OCT-2004 (Rel. 45, Created)
DT   25-OCT-2004 (Rel. 45, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Methyl-CpG binding protein 4 (EC 3.2.2.-) (Methyl-CpG binding domain
DE   protein 4) (Mismatch-specific DNA N-glycosylase).
GN   Name=Mbd4;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=98449942; PubMed=9774669;
RA   Hendrich B., Bird A.;
RT   "Identification and characterization of a family of mammalian methyl-
RT   CpG binding proteins.";
RL   Mol. Cell. Biol. 18:6538-6547(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=99373255; PubMed=10441743;
RA   Hendrich B., Abbott C., McQueen H., Chambers D., Cross S.H., Bird A.;
RT   "Genomic structure and chromosomal mapping of the murine and human
RT   mbd1, mbd2, mbd3, and mbd4 genes.";
RL   Mamm. Genome 10:906-912(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 411-554.
RX   PubMed=12456671; DOI=10.1074/jbc.M210884200;
RA   Wu P., Qiu C., Sohail A., Zhang X., Bhagwat A.S., Cheng X.;
RT   "Mismatch repair in methylated DNA. Structure and activity of the
RT   mismatch-specific thymine glycosylase domain of methyl-CpG-binding
RT   protein MBD4.";
RL   J. Biol. Chem. 278:5285-5291(2003).
CC   -!- FUNCTION: Mismatch-specific DNA N-glycosylase involved in DNA
CC       repair. Has thymine glycosylase activity and is specific for G:T
CC       mismatches within methylated and unmethylated CpG sites. Can also
CC       remove uracil or 5-fluorouracil in G:U mismatches. Has no lyase
CC       activity. Was first identified as methyl-CpG-binding protein.
CC   -!- SUBUNIT: Interacts with MLH1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear, in discrete foci.
CC   -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072249; AAC68878.1; -.
DR   EMBL; AF120996; AAD56595.1; -.
DR   EMBL; BC024812; AAH24812.1; -.
DR   PDB; 1NGN; X-ray; A=400-554.
DR   MGD; MGI:1333850; Mbd4.
DR   InterPro; IPR003265; Endo_3c.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF00730; HhH-GPD; 1.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
DR   PROSITE; PS50982; MBD; 1.
KW   3D-structure; DNA repair; DNA-binding; Hydrolase; Nuclear protein.
FT   DOMAIN       63    135       MBD.
FT   DOMAIN      435    498       HhH-GPD.
FT   ACT_SITE    534    534       By similarity.
FT   CONFLICT    129    129       N -> D (in Ref. 3).
SQ   SEQUENCE   554 AA;  62577 MW;  792D37CB180291F5 CRC64;

  Query Match             59.6%;  Score 1821.5;  DB 1;  Length 554;
  Best Local Similarity   66.2%;  Pred. No. 7.6e-103;
  Matches  384;  Conservative   49;  Mismatches  116;  Indels   31;  Gaps   11;

Qy          6 LESLSLGD---RGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSECN 62
              :|| :|||   ||         | ||||||  |  |||:|: |  ||||  : ::|  ||| :
Db          1 MESPNLGDNRVRG--------ESLVPDPPWDRCKEDIAVGLGGVGEDGKDLVI--SSERS 50

Qy         63 PLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKS 122
               |||||  ||    :|  ||  | |||||||||||||| |||||:|||||||||||||||||
Db         51 SLLQEPTAST-LSSTTATEGHKPVPCGWERVVKQRLSGKTAGKFDVYFISPQGLKFRSKR 109

Qy        123 SLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRS 182
              ||||||  |||||  ||||||:|||| |  |   ||  |:|||||   |::: |  ||:|||
Db        110 SLANYLLKNGETFLKPEDFNFTVLPKGSINPGYKHQSLAALTSLQPNETDVSKQNLKTRS 169

Qy        183 KCKKDVFMPPSSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIP 242
              | | ||   || :||  || ||||  |  |||:|   : ||:  | || | |||:|||
Db        170 KWKTDVLPLPSGTSESPESSGLSNSNSACLLLREHRDIQDVDSEKRRKSKRKVTVLKGTA 229

Qy        243 IKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLSV 302
              :|||: ||||     | : || ||  |  |: | | |:|||:||:| :|| |  ||: :
Db        230 SQKTKQKCRKSLLESTQRNRKRASVVQKVGADRELVPQESQLNRTLCPADACA-RETVGL 288

Qy        303 TSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGTKV 362
              |       |:| | | : |  | |||  ||| | : :    : | ||||||| :|
Db        289 AGE--------EKSPSPGLDLCFIQVTSGTTNKFHSTEAAGEANR-EQTFLESEEIRSK- 338

Qy        363 EVVERK--EHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
                :||   |||| :|: ||||  :||  :| || |  ||||:||||:|:|||||||||
```

```
Db        339 --GDRKGEAHLHTGVLQDGSEMP-SCSQAKKHFTSE-TFQEDSIPRTQVEKRKTSLYFSS 394

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              |||||||||||||:||||||||||||||| ||||||||||||||||||||||||||||:
Db        395 KYNKEALSPPRRKSFKKWTPPRSPFNLVQEILFHDPWKLLIATIFLNRTSGKMAIPVLWE 454

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              |||||||||||| ||||||||||||||||||||:||||||||||:||||||||||
Db        455 FLEKYPSAEVARAADWRDVSELLKPLGLYDLRAKTIIKFSDEYLTKQWRYPIELHGIGKY 514

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              |||||||||||||||||||||||||||||||||||||||
Db        515 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 554


RESULT 3
Q9I9F1
ID   Q9I9F1      PRELIMINARY;      PRT;   416 AA.
AC   Q9I9F1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   5-methylcytosine G/T mismatch-specific DNA glycosylase.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20512562; PubMed=11058112; DOI=10.1093/nar/28.21.4157;
RA   Zhu B., Zheng Y., Angliker H., Schwarz S., Thiry S., Siegmann M.,
RA   Jost J.-P.;
RT   "5-Methylcytosine DNA glycosylase activity is also present in the
RT   human MBD4 (G/T mismatch glycosylase) and in a related avian
RT   sequence.";
RL   Nucleic Acids Res. 28:4157-4165(2000).
DR   EMBL; AF257107; AAF68981.1; -.
DR   HSSP; Q9Z2D7; 1NGN.
DR   InterPro; IPR011257; DNA_glycsylse.
SQ   SEQUENCE   416 AA;  45454 MW;  A3F70C6FF2133F2A CRC64;

  Query Match             28.8%;  Score 880.5;  DB 2;  Length 416;
  Best Local Similarity   78.2%;  Pred. No. 1.2e-45;
  Matches  161;  Conservative   20;  Mismatches   21;  Indels    4;  Gaps    2;

Qy        378 RGSEMDNNCS--PTRKDFTGEKIFQ--EDTIPRTQIERRKTSLYFSSKYNKEALSPPRRK 433
              | |  | : |    :| ||  :  :  |:: ||||::||||| ||||||:||||||||||
Db        210 RDSAADGDVSWPSDKKSFTAVQAPRGTEESAPRTQVDRRKTSPYFSSKYSKEALSPPRRK 269

Qy        434 AFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVART 493
              ||:|||||||||||||||||||||||||||||:||||||||||:|| |||| |||||
Db        270 AFRKWTPPRSPFNLVQETLFHDPWKLLIATIFLNKTSGKMAIPVLWEFLRKYPSPEVART 329

Qy        494 ADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEW 553
              |||:::|||||||||| |||||||:||||||| ||||||||||||||||||||||||||
Db        330 ADWKEMSELLRPLGLYALRAKTIIKFSDEYLNKQWKYPIELHGIGKYGNDSYRIFCVNEW 389

Qy        554 KQVHPEDHKLNKYHDWLWENHEKLSL 579
              |:| |:||||| || ||||||||||:
Db        390 KEVQPQDHKLNIYHAWLWENHEKLSV 415


RESULT 4
Q9SFC1
ID   Q9SFC1      PRELIMINARY;      PRT;   419 AA.
AC   Q9SFC1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F17A17.27 protein.
GN   Name=F17A17.27;
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lin X., Kaul S., Town C.D., Benito M.-I., Creasy T.H., Haas B.,
RA   Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA   Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC013483; AAF21203.1; -.
DR   HSSP; Q9Z2D7; 1NGN.
DR   GO; GO:0006284; P:base-excision repair; IEA.
DR   InterPro; IPR011257; DNA_glycsylse.
DR   InterPro; IPR003265; Endo_3c.
DR   Pfam; PF00730; HhH-GPD; 1.
SQ   SEQUENCE   419 AA;  47957 MW;  BAFF0BCA5A710C95 CRC64;

  Query Match             11.3%;  Score 344;  DB 2;  Length 419;
  Best Local Similarity   25.4%;  Pred. No. 6.1e-13;
  Matches   99;  Conservative   57;  Mismatches  146;  Indels   88;  Gaps   11;

Qy        205 SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPI-------KKTKKGCRKSCSGF 257
              ||  |  :   :|: | | :  :    :  | : :  |        :::|:||
Db         88 SNLVSPQIADDDDDSVSDSHIERQECSEFHVEVRRVSPYFQGSTVSQQSKEGC------- 140

Qy        258 VQSDSKRESVCNKADAES--------EPVAQKSQLDRTVCISDAGACGET----LSVTSE 305
                ||  :|||:|               |  | | : :  | ||  :  ::      :|:
Db        141 ---DS--DSVCSKEGCSKVQAKVPRVSPYFQASTISQ--CDSDIVSSSQSGRNYRKGSSK 193

Qy        306 ENSLVKK-----KERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
                |::     :| ::|   |    |   |: | |
Db        194 RQVKVRRVSPYFQESTVSEQPN----QAPKGLRNYF---------------------- 225

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              :||:   : | | ::            ||            |   : ::   :|
Db        226 --KVVKVSRYFHADGIQVNESQKEKSRNVRK--------TPIVSPVLSLSQKTDDVYL-- 273

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
                      |:    | ||||| ||:||  :||||::|:  : ||:|||     |:
Db        274 ----------RKTPDNTWVPPRSPCNLLQEDHWHDPWRVLVICMLLNKTSGAQTRGVISD 323

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
                    |: |      ::  |:|||||   | | | : | ||| : | :  :|||:|||
Db        324 LFGLCTDAKTATEVKEEEIENLIKPLGLQKKRTKMIQRLSLEYLQESWTHVTQLHGVGKY 383

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWL 570
              |:| |||   | :| | || || | |:|
Db        384 AADAYAIFCNGNWDRVKPNDHMLNYYWDYL 413


RESULT 5
Q84WT3
ID   Q84WT3      PRELIMINARY;      PRT;   407 AA.
AC   Q84WT3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein At3g07930.
GN   Name=At3g07930;
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA   Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,
RA   Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA   Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA   Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
```

```
RA   Theologis A.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT002799; AAO22623.1; -.
DR   HSSP; Q9Z2D7; 1NGN.
DR   GO; GO:0006284; P:base-excision repair; IEA.
DR   InterPro; IPR011257; DNA_glycsylse.
DR   InterPro; IPR003265; Endo_3c.
DR   Pfam; PF00730; HhH-GPD; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   407 AA;  46941 MW;  DD758CD8862EF54F CRC64;

  Query Match             11.1%;  Score 338;  DB 2;  Length 407;
  Best Local Similarity   26.6%;  Pred. No. 1.4e-12;
  Matches  102;  Conservative   51;  Mismatches  155;  Indels   76;  Gaps   11;

Qy        251 RKSCSGFVQ-SDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSL 309
Db         30 RRPDSDFIEVSDENRSFALFKEDDEK---------NRDLGLVDDGSTNLVLQCHDDGCSL  80

Qy        310 VKKKERSLS---SGSNFCSEQK----------TSGIINKFCSAKDSEHNEKYEDTFLESE 356
Db         81 EKDNSNSLDDLFSGFVYKGVRRRKRDDFGSITTSNLVSPQIADDD---DDSVSDSHIERQ 137

Qy        357 EIGTKVEVVERK---------EHLHTDILKRGSEMDN-NCSPTRKDFTGEKI---FQEDT 403
Db        138 ECSKVQAKVPRVSPYFQASTISQCDSDIVSSSQSGRNYRKGSSKRQVKARRVSPYFQEST 197

Qy        404 IP----------RTQIERRKTSLYF--------------SSKYNKEALSPP--------- 430
Db        198 VSEQPNQAPKGLRNYFKVVKVSRYFHADGIQVNESQKEKSRNVRKTPIVSPVLSLSQKTD 257

Qy        431 ----RRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYP 486
Db        258 DVYLRKTPDNTWVPPRSPCNLLQEDHWHDPWRVLVICMLLNKTSGAQTRGVISDLFGLCT 317

Qy        487 SAEVARTADWRDVSELLKPLGLGYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYR 546
Db        318 DAKTATEVKEEEIENLIKPLGLQKKRTKMIQRLSLEYLQESWTHVTQLHGVGKYAADAYA 377

Qy        547 IFCVNEWKQVHPEDHKLNKYHDWL 570
Db        378 IFCNGNWDRVKPNDHMLNYYWDYL 401


RESULT 6
Q7T2T7
ID   Q7T2T7      PRELIMINARY;      PRT;   524 AA.
AC   Q7T2T7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Methyl-cytosine binding protein 2.
GN   Name=mecp2;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Coverdale L.E., Martin C.C.;
RL   Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY298900; AAP57248.2; -.
DR   HSSP; Q9UNZ9; 1IG4.
DR   ZFIN; ZDB-GENE-030131-7190; mecp2.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   524 AA;  57152 MW;  B8593B4BB84DC2D1 CRC64;

  Query Match              7.8%;  Score 238.5;  DB 2;  Length 524;
  Best Local Similarity   24.4%;  Pred. No. 2.2e-06;
  Matches  106;  Conservative   57;  Mismatches  159;  Indels  113;  Gaps   18;

Qy         27 VPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGA---TAGTECR 83
Db         49 VPPPPSLFTQRDVGQQAE-AGKSE-----------PI--DPEVGAALSAPESSASAKQR 93

Qy         84 KSV------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHK 130
Db         94 RSVIRDRGPMYEDPSLPQGWTRKLKQRKSGRSAGKFDVYLINPEGKAFRSKVELMAYFQK 153

Qy        131 NGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFM 190
Db        154 VGDTITDPNDFDFTV-TGRGSPSR----------------------REKRPPKKPKMV 188

Qy        191 PPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTI------------- 237
Db        189 KPS-------GRGRGRPKGSGKVRQATEGV--AVKRVIEKSPGKLLVKMPFVAPKTEPGA 239

Qy        238 -LKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGAC 296
Db        240 PLGQAPVAKARRGRKR------KSEQDPPSTPKKRGRKPATVSQS-----TVGTGSAAAY 288

Qy        297 GETLSVTSE-------ENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYE 349
Db        289 AAAAILTAEAKKKALKESSAKPVQERALP-----IKKRKTRETLEEL-EASTTSATETFE 342

Qy        350 DTFLESEEIGTKVEV-VERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQ 408
Db        343 KRLTASTVTPTGEEAETGQKPHKHPS--RKHKEADPGSSSSGTTASG-------VAPKSH 393

Qy        409 IERRKTSLYFSSKYN 423
Db        394 KKRDQRGQHFKHHHH 408


RESULT 7
Q9YGC6
ID   Q9YGC6      PRELIMINARY;      PRT;   467 AA.
AC   Q9YGC6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Methyl-CpG-binding protein MeCP2 (Methyl-CpG-binding protein 2).
GN   Name=MECP2;
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98282101; PubMed=9620779; DOI=10.1038/561;
RA   Jones P.L., Veenstra G.J.C., Wade P.A., Vermaak D., Kass S.U.,
RA   Landsberger N., Strouboulis J., Wolffe A.P.;
RT   "Methylated DNA and MeCP2 recruit histone deacetylase to repress
RT   transcription.";
RL   Nat. Genet. 19:187-191(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kass S.U., Strouboulis J., Wolffe A.P.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF106951; AAD03736.1; -.
DR   EMBL; AF051768; AAD02651.1; -.
DR   HSSP; P51608; 1QK9.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000637; A+T_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00384; AT_hook; 2.
DR   SMART; SM00391; MBD; 1.
```

```
SQ   SEQUENCE   467 AA;  51757 MW;  5D3A719A59E560BC CRC64;

  Query Match             7.2%;  Score 220.5;  DB 2;  Length 467;
  Best Local Similarity   25.7%;  Pred. No. 2.3e-05;
  Matches   98;  Conservative   60;  Mismatches  157;  Indels   67;  Gaps   15;

Qy         22 SSERLVPDPPNDLRKEDVAMELER---VGEDEEQMMIKRS--SECNPLLQEPIASAQFGA 76
              |||   |  | |  | |::   |    | |       :||   :  |: ::|
Db         49 SSEH-QPGEPADEGKADMSESAEENLAVPESSASPKQRRSVIRDRGPMYEDP-------- 99

Qy         77 TAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSL 136
                ::| || | :|||  |::||:|||| |:| |  ||||  |  |  | |:|||
Db        100 --------TLPEGWTRKLKQRKSGRSAGKFDVYLINPNGKAFRSKVELIAYFQKVGDTSL 151

Qy        137 KPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDV--FMPPSS 194
              | |||||| : ||  || :         : ::  |: : | | : |  :    ||
Db        152 DPNDFDFTV-TGRGSPSRREQ------KQPKKPKAPKSSVSGRGRGRPKGSIKKVKPPVK 204

Qy        195 SSELQESRGLSNFTSTHLLLK-----EDEGVDDVNFRKV-------RKPKGKVTILKGIP 242
              |  :|  | :   :  ||:|       |  |    ::|       || | : |      |
Db        205 SEGVQVKRVIEK-SPGKLLVKMPYSGTKEASDATTSQQVLVIKRGGRKRKSE-TDPSAAP 262

Qy        243 IKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLSV 302
              |: :|    | :      :|::::    ::  :|: :       ||         ||:||
Db        263 KKRGRKPSNVSLAAAAAAEAAKKKAI---KESSIKPLLE------TVLPIKKRKTRETISV 313

Qy        303 TSEE-------NSLVKK------KERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYE 349
              ::         :::|        :|  | |    | :  :|:  |        |:  :
Db        314 DVKDTIKPEPLTPVIEKVMKGQNPAKSPESRSTEGSPKIKTGLPKKELQQHHHHHHHHHH 373

Qy        350 DTFLESEEIGTKVEVVERKEHL 371
                ||:   |  |    |:::
Db        374 HHHSESKASATSPEPETSKDNI 395


RESULT 8
Q7SCQ2
ID   Q7SCQ2      PRELIMINARY;      PRT;   682 AA.
AC   Q7SCQ2;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=NCU09815.1;
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OR74A;
RA   Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA   Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA   Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA   Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA   Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA   Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA   Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA   Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA   Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA   Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A.,
RA   DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA   Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA   Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA   Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
RT   "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL   Nature 0:0-0(2003).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABX01000087; EAA34519.1; -.
DR   GO; GO:0006284; P:base-excision repair; IEA.
DR   InterPro; IPR011257; DNA_glycsylse.
DR   InterPro; IPR003265; Endo_3c.
DR   Pfam; PF00730; HhH-GPD; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   682 AA;  75928 MW;  4FDCCAA26102E8E4 CRC64;

  Query Match             7.0%;  Score 213;  DB 2;  Length 682;
  Best Local Similarity   20.8%;  Pred. No. 0.00011;
  Matches   93;  Conservative   62;  Mismatches  155;  Indels  138;  Gaps   14;

Qy        240 GIPIKKTKKGCRKSCSGF-VQSDSKRES-------------VCNKADAESEPVAQKSQL 284
              |  :|||     :  | |  | |  : |                    ||    ||:::: :
Db        164 GKRVKKTTGVMGRGVSRFWAQDDPAKTSRQVTRGRGETGVSGAVRKAMNGHEPISEQGK- 222

Qy        285 DRTVCISDAGACGETLSVTSEENSLVKKKE-RSLSSGSNFCSEQKTSGIINKFCSAKDSE 343
                          ::: :   |:| :    |   |    |: | | : :|    | |:
Db        223 --------------MALPTRVMSVVHRSAGPSQGEGE---SKAKKSRLTDK-TKLKHSK 263

Qy        344 HNEK--YEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNC-SPTRKDFTGEKIFQ 400
                ::  | ||:     : :     |  :    |:   |:   : |          |||: ::
Db        264 FPDEPSYADTWSPQRSLASHSSPNENLDVALGDVSVAGTSSFDECVLAANVRITGERTYK 323

Qy        401 EDTI---PRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPP----------RSPFN 446
                    |    | |   |     :: || ::   :   ||             | |
Db        324 SPFFSEPPPWPISTSKASPPKSKTSTSKSRSPTKKS--RNPHPPGISCLPIAPISASRFG 381

Qy        447 LVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPL 506
              |:|| :  ||::|||  || :  | ||||:  : :: :|: |   :||  ::  |::||
Db        382 LIQEEVAADPFRLLIAVTFLIKVRGTMAIPLFRQLMDLFPTPEALASADPSEIINLIRPL 441

Qy        507 GLYDLRAKTIVKFS----------DEYLTKQWKYPIELHGI----------------- 537
              ||   |   | |::                |  | :  | :  |:
Db        442 GLSVNRCSVIQKYARMFIECPPCKEKRYGVKNYPRPGDARGVKVGQEFTGEPDDFHIGKA 501

Qy        538 ---------------------------------GKYGNDSYRIFCVN---------- 551
                                                  | |  ||:|||| :
Db        502 SQAEFDDDDDRINAIKFAKEHAIGCAWEIGHLTQGPYALDSWRIFCRDKLLGRADDWKGK 561

Qy        552 --------EWKQVHPEDHKLNKYHDWLW 571
                     || :| |:| :|  |  |:|
Db        562 GRHPEFQPEWMRVLPQDKELRAYLRWMW 589


RESULT 9
MEC2_HUMAN
ID   MEC2_HUMAN     STANDARD;      PRT;   486 AA.
AC   P51608; O15233;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN   Name=MECP2;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kudo S., Fukuda M.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Thiesen J., Straetling W.H.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97130625; PubMed=8976388;
RA   Vilain A., Apiou F., Vogt N., Dutrillaux B., Malfoy B.;
RT   "Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to
RT   human chromosome band Xq28 by in situ hybridization.";
RL   Cytogenet. Cell Genet. 74:293-294(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
```

RA Reichwald K., Rosenthal A., Kioschis P., Platzer M.;
RT "Mapping and sequence analysis of the human MeCP2 locus.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=99299240; PubMed=10369871; DOI=10.1093/hmg/8.7.1253;
RA Coy J.F., Sedlacek Z., Baechner D., Delius H., Poustka A.;
RT "A complex pattern of evolutionary conservation and alternative
RT polyadenylation within the long 3'-untranslated region of the methyl-
RT CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene
RT expression.";
RL Hum. Mol. Genet. 8:1253-1262(1999).
RN [6]
RP SEQUENCE FROM N.A.
RA Reichwald K., Thiesen J., Wiehe T., Weitzel J., Straetling W.H.,
RA Kioschis P., Rosenthal A., Platzer M.;
RT "Comparative sequence analysis of the MECP2-locus in human and mouse
RT reveals new untranslated regions.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [8]
RP SEQUENCE OF 10-486 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96327611; PubMed=8672133;
RA D'Esposito M., Quaderi N.A., Ciccodicola A., Bruni P., Esposito T.,
RA D'Urso M., Brown S.D.M.;
RT "Isolation, physical mapping, and Northern analysis of the X-linked
RT human gene encoding methyl CpG-binding protein, MECP2.";
RL Mamm. Genome 7:533-535(1996).
RN [9]
RP SEQUENCE OF 10-486 FROM N.A.
RA Reichwald K., Bauer D., Brenner V., Drescher B., Coy J.F.,
RA Kioschis P., Korn B., Nyakatura G., Platzer M., Poustka A.,
RA Sandoval N., Rosenthal A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN [10]
RP REVIEW ON VARIANTS.
RX PubMed=12872250; DOI=10.1002/humu.10243;
RA Miltenberger-Miltenyi G., Laccone F.;
RT "Mutations and polymorphisms in the human methyl CpG-binding protein
RT MECP2.";
RL Hum. Mutat. 22:107-115(2003).
RN [11]
RP VARIANTS RTT TRP-106; CYS-133; SER-155; MET-158 AND CYS-306, AND
RP VARIANT LYS-397.
RX PubMed=10577905;
RA Wan M., Lee S.S.J., Zhang X., Houwink-Manville I., Song H.-R.,
RA Amir R.E., Budden S., Naidu S., Pereira J.L.P., Lo I.F.M.,
RA Zoghbi H.Y., Schanen N.C., Francke U.;
RT "Rett syndrome and beyond: recurrent spontaneous and familial MECP2
RT mutations at CpG hotspots.";
RL Am. J. Hum. Genet. 65:1520-1529(1999).
RN [12]
RP VARIANTS RTT TRP-106; CYS-133; SER-155 AND MET-158.
RX MEDLINE=99438392; PubMed=10508514; DOI=10.1038/13810;
RA Amir R.E., Van den Veyver I.B., Wan M., Tran C.Q., Francke U.,
RA Zoghbi H.Y.;
RT "Rett syndrome is caused by mutations in X-linked MECP2, encoding
RT methyl-CpG-binding protein 2.";
RL Nat. Genet. 23:185-188(1999).
RN [13]
RP INVOLVEMENT IN X-LINKED MENTAL RETARDATION WITH PROGRESSIVE
RP SPASTICITY.
RX PubMed=10986043;
RA Meloni I., Bruttini M., Longo I., Mari F., Rizzolio F., D'Adamo P.,
RA Denvriendt K., Fryns J.-P., Toniolo D., Renieri A.;
RT "A mutation in the Rett syndrome gene, MECP2, causes X-linked mental
RT retardation and progressive spasticity in males.";
RL Am. J. Hum. Genet. 67:982-985(2000).
RN [14]
RP VARIANTS RTT VAL-100; GLN-106; TRP-106; CYS-133; ARG-152; SER-155;
RP MET-158; ARG-305; CYS-306 AND HIS-306, AND VARIANTS CYS-86; MET-203;
RP PRO-287; ALA-291; LYS-397; ILE-412 AND THR-444.
RX PubMed=11055898;
RA Buyse I.M., Fang P., Hoon K.T., Amir R.E., Zoghbi H.Y., Roa B.B.;
RT "Diagnostic testing for Rett syndrome by DHPLC and direct sequencing
RT analysis of the MECP2 gene: identification of several novel mutations
RT and polymorphisms.";
RL Am. J. Hum. Genet. 67:1428-1436(2000).
RN [15]
RP VARIANT MRX16 VAL-140, AND VARIANT MET-203.
RX MEDLINE=20465115; PubMed=11007980; DOI=10.1016/S0014-5793(00)01994-3;
RA Orrico A., Lam C., Galli L., Dotti M.T., Hayek G., Tong S.F.,
RA Poon P.M., Zappella M., Federico A., Sorrentino V.;
RT "MECP2 mutation in male patients with non-specific X-linked mental
RT retardation.";
RL FEBS Lett. 481:285-288(2000).
RN [16]
RP VARIANTS RTT LEU-101; HIS-101; THR-101; TRP-106; CYS-133; CYS-134;
RP ARG-152; MET-158; ARG-225; LEU-302; CYS-306 AND HIS-306, AND VARIANTS
RP LEU-229 AND THR-439.
RX PubMed=10767337; DOI=10.1093/hmg/9.7.1119;
RA Cheadle J.P., Gill H., Fleming N., Maynard J., Kerr A., Leonard H.,
RA Krawczak M., Cooper D.N., Lynch S., Thomas N., Hughes H., Hulten M.,
RA Ravine D., Sampson J.R., Clarke A.;
RT "Long-read sequence analysis of the MECP2 gene in Rett syndrome
RT patients: correlation of disease severity with mutation type and
RT location.";
RL Hum. Mol. Genet. 9:1119-1129(2000).
RN [17]
RP VARIANTS RTT GLN-106; MET-158; ARG-302; CYS-306 AND ALA-322.
RX PubMed=10814719; DOI=10.1093/hmg/9.9.1377;
RA Bienvenu T., Carrie A., de Roux N., Vinet M.-C., Jonveaux P.,
RA Couvert P., Villard L., Arzimanoglou A., Beldjord C., Fontes M.,
RA Tardieu M., Chelly J.;
RT "MECP2 mutations account for most cases of typical forms of Rett
RT syndrome.";
RL Hum. Mol. Genet. 9:1377-1384(2000).
RN [18]
RP VARIANTS RTT MET-158; HIS-302 AND CYS-306, AND VARIANTS VAL-201;
RP ALA-232; LEU-251 AND SER-376.
RX PubMed=10944854;
RA Amano K., Nomura Y., Segawa M., Yamakawa K.;
RT "Mutational analysis of the MECP2 gene in Japanese patients with Rett
RT syndrome.";
RL J. Hum. Genet. 45:231-236(2000).
RN [19]
RP VARIANTS RTT TRP-106; PHE-124; CYS-133; CYS-134; ARG-152; MET-158 AND
RP CYS-306.
RX MEDLINE=20439334; PubMed=10991688;
RA Obata K., Matsuishi T., Yamashita Y., Fukuda T., Kuwajima K.,
RA Horiuchi I., Nagamitsu S., Iwanaga R., Kimura A., Omori I., Endo S.,
RA Mori K., Kondo I.;

```
RT   "Mutation analysis of the methyl-CpG binding protein 2 gene (MECP2) in
RT   patients with Rett syndrome.";
RL   J. Med. Genet. 37:608-610(2000).
RN   [20]
RP   VARIANTS RTT ARG-101; TRP-106; MET-158 AND CYS-306, AND VARIANT
RP   LYS-397.
RX   MEDLINE=20439335; PubMed=10991689;
RA   Hampson K., Woods C.G., Latif F., Webb T.;
RT   "Mutations in the MECP2 gene in a cohort of girls with Rett
RT   syndrome.";
RL   J. Med. Genet. 37:610-612(2000).
RN   [21]
RP   VARIANT RTT HIS-133.
RX   PubMed=11706982; DOI=10.1002/ana.1272;
RA   Armstrong J., Poo P., Pineda M., Aibar E., Gean E., Catala V.,
RA   Monros E.;
RT   "Classic Rett syndrome in a boy as a result of somatic mosaicism for a
RT   MECP2 mutation.";
RL   Ann. Neurol. 50:692-692(2001).
RN   [22]
RP   VARIANTS RTT TRP-106; CYS-134; ARG-152; MET-158; ALA-302; CYS-306 AND
RP   ALA-322, AND VARIANTS VAL-201 AND LYS-397.
RX   PubMed=11738883; DOI=10.1016/S0387-7604(01)00342-4;
RA   Giunti L., Pelagatti S., Lazzerini V., Guarducci S., Lapi E.,
RA   Coviello S., Cecconi A., Ombroni L., Andreucci E., Sani I.,

  Query Match             6.8%;  Score 208.5;  DB 1;  Length 486;
  Best Local Similarity   23.9%;  Pred. No. 0.00013;
  Matches  104;  Conservative   53;  Mismatches  149;  Indels  129;  Gaps   17;

Qy         37 EDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG----------------ATAGT 80
              :|  :: ::| :|:::    :     |         |: |                |:|
Db         22 KDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASP 81

Qy         81 ECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANY 127
              : |:|:              | || | :|||  |::||::||| |:|||  ||||  |  |
Db         82 KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAY 141

Qy        128 LHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKD 187
               | |:||| | |||||| : ||  ||                       | :   ||
Db        142 FEKVGDTSLDPNDFDFTV-TGRGSPSR----------------------REQKPPKK- 175

Qy        188 VFMPPSSSSELQESRGL---SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIK 244
              | |         ||        :|       |||     | : |  ||  :|  :|
Db        176 ----PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF- 226

Qy        245 KTKKGCRKSCSGFVQSDS----KRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              :|  | :    |   |       ||      ||:|: : : :|                | :
Db        227 QTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKR----------GRKPGSVV 276

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              :  : |      || |:       | |:|   | |           |  :|
Db        277 AAAAAEAKKKAVKESSIR------SVQETVLPIKK----------RKTRET--------V 312

Qy        361 KVEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYF 418
              :|| |  : |     |  | |  :    || ||         ::: |: :
Db        313 SIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSPKGR---------- 354

Qy        419 SSKYNKEALSPPRRK 433
                :  | |||:::
Db        355 ----SSSASSPPKKE 365


RESULT 10
Q6QHH9
ID   Q6QHH9      PRELIMINARY;      PRT;   498 AA.
AC   Q6QHH9;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Methyl CpG binding protein 2 isoform B.
GN   Name=MECP2;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=15034579; DOI=10.1038/ng1327;
RA   Mnatzakanian G.N., Lohi H., Munteanu I., Alfred S.E., Yamada T.,
RA   MacLeod P.J., Jones J.R., Scherer S.W., Schanen N.C., Friez M.J.,
RA   Vincent J.B., Minassian B.A.;
RT   "A previously unidentified MECP2 open reading frame defines a new
RT   protein isoform relevant to Rett syndrome.";
RL   Nat. Genet. 36:339-341(2004).
DR   EMBL; AY541280; AAS55455.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000637; A+T_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00391; MBD; 1.
SQ   SEQUENCE   498 AA;  53323 MW;  443ECB3D5EA4DAB8 CRC64;

  Query Match             6.8%;  Score 208.5;  DB 2;  Length 498;
  Best Local Similarity   23.9%;  Pred. No. 0.00014;
  Matches  104;  Conservative   53;  Mismatches  149;  Indels  129;  Gaps   17;

Qy         37 EDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG----------------ATAGT 80
              :|  :: ::| :|:::    :     |         |: |                |:|
Db         34 KDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASP 93

Qy         81 ECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANY 127
              : |:|:              | || | :|||  |::||::||| |:|||  ||||  |  |
Db         94 KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAY 153

Qy        128 LHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKD 187
               | |:||| | |||||| : ||  ||                       | :   ||
Db        154 FEKVGDTSLDPNDFDFTV-TGRGSPSR----------------------REQKPPKK- 187

Qy        188 VFMPPSSSSELQESRGL---SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIK 244
              | |         ||        :|       |||     | : |  ||  :|  :|
Db        188 ----PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF- 238

Qy        245 KTKKGCRKSCSGFVQSDS----KRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              :|  | :    |   |       ||      ||:|: : : :|                | :
Db        239 QTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKR----------GRKPGSVV 288

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              :  : |      || |:       | |:|   | |           |  :|
Db        289 AAAAAEAKKKAVKESSIR------SVQETVLPIKK----------RKTRET--------V 324

Qy        361 KVEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYF 418
              :|| |  : |     |  | |  :    || ||         ::: |: :
Db        325 SIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSPKGR---------- 366

Qy        419 SSKYNKEALSPPRRK 433
                :  | |||:::
Db        367 ----SSSASSPPKKE 377


RESULT 11
Q7Z384
ID   Q7Z384      PRELIMINARY;      PRT;   516 AA.
AC   Q7Z384;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein DKFZp686A24160 (Fragment).
GN   Name=DKFZp686A24160;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Human colon endothel primary cell culture;
RA   Bloecker H., Boecher M., Mewes H.W., Weil B., Amid C., Osanger A.,
RA   Fobo G., Han M., Wiemann S.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX538060; CAD97991.1; -.
DR   HSSP; Q9UNZ9; 1IG4.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000637; A+T_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00384; AT_hook; 2.
DR   SMART; SM00391; MBD; 1.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   516 AA;  55204 MW;  27CD37B9164176B0 CRC64;

  Query Match             6.8%;  Score 208.5;  DB 2;  Length 516;
  Best Local Similarity   23.9%;  Pred. No. 0.00014;
  Matches  104;  Conservative   53;  Mismatches  149;  Indels  129;  Gaps   17;

Qy         37 EDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG----------------ATAGT 80
              :|  :: ::| :|:::    :     |         |: |                |:|
Db         52 KDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASP 111

Qy         81 ECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANY 127
              : |:|:             | || | :|||  |::||::||| |:||| ||||  |  |
Db        112 KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAY 171

Qy        128 LHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKD 187
               `| |:||| | |||||| : ||  ||                        | :   ||
Db        172 FEKVGDTSLDPNDFDFTV-TGRGSPSR-----------------------REQKPPKK- 205

Qy        188 VFMPPSSSSELQESRGL---SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIK 244
               ,  | |        ||        :|       |||     | : |  ||  :|  :|
Db        206 ----PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF- 256

Qy        245 KTKKGCRKSCSGFVQSDS----KRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              :|  | :   |  |       ||     ||:|: : : :|                |  :
Db        257 QTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKR----------GRKPGSVV 306

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              :  : |       || |:       | |:|   | |           |  :|
Db        307 AAAAAEAKKKAVKESSIR------SVQETVLPIKK----------RKTRET--------V 342

Qy        361 KVEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYF 418
              :|| |  : |     |  | |  :    || ||          ::: |: :
Db        343 SIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSPKGR---------- 384

Qy        419 SSKYNKEALSPPRRK 433
                :  | |||:::
Db        385 ----SSSASSPPKKE 395


RESULT 12
MEC2_MACFA
ID   MEC2_MACFA     STANDARD;      PRT;   486 AA.
AC   Q95LG8;
DT   25-OCT-2004 (Rel. 45, Created)
DT   25-OCT-2004 (Rel. 45, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN   Name=MECP2;
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Muramatsu S.;
RT   "Excessive hand-wringing in a MPTP-treated monkey.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Chromosomal protein that binds to methylated DNA. It can
CC       bind specifically to a single methyl-CpG pair. It is not
CC       influenced by sequences flanking the methyl-CpGs. Mediates
CC       transcriptional repression through interaction with histone
CC       deacetylase and the corepressor SIN3A.
CC   -!- SUBUNIT: Interacts with FNBP3 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Colocalized with methyl-CpG in the
CC       genome (By similarity).
CC   -!- SIMILARITY: Contains 2 A.T hook DNA-binding repeats.
CC   -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF295597; AAK97131.1; -.
DR   HSSP; P51608; 1QK9.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00384; AT_hook; 2.
DR   SMART; SM00391; MBD; 1.
DR   PROSITE; PS50982; MBD; 1.
KW   DNA-binding; Nuclear protein; Repeat; Repressor;
KW   Transcription regulation.
FT   DOMAIN       90     162       MBD.
FT   DNA_BIND     185    197       A.T hook 1.
FT   DNA_BIND     265    277       A.T hook 2.
FT   DOMAIN       366    372       His-rich.
FT   DOMAIN       376    405       Pro-rich.
SQ   SEQUENCE   486 AA;  52426 MW;  3471B61D90D92A7D CRC64;

  Query Match             6.8%;  Score 207.5;  DB 1;  Length 486;
  Best Local Similarity   23.9%;  Pred. No. 0.00015;
  Matches  104;  Conservative   53;  Mismatches  149;  Indels  129;  Gaps   17;

Qy         37 EDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG----------------ATAGT 80
              :|  :: ::| :|:::    :     |         |: |                |:|
Db         22 KDKPLKFKKVKKDKKEDKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASP 81

Qy         81 ECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANY 127
              : |:|:             | || | :|||  |::||::||| |:||| ||||  |  |
Db         82 KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAY 141

Qy        128 LHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKD 187
               | |:||| | |||||| : ||  ||                        | :   ||
Db        142 FEKVGDTSLDPNDFDFTV-TGRGSPSR-----------------------REQKPPKK- 175

Qy        188 VFMPPSSSSELQESRGL---SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIK 244
               | |        ||        :|       |||     | : |  ||  :|  :|
Db        176 ----PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF- 226

Qy        245 KTKKGCRKSCSGFVQSDS----KRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              :|  | :   |  |       ||     ||:|: : : :|                |  :
Db        227 QTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKR----------GRKPGSVV 276

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              :  : |       || |:       | |:|   | |           |  :|
Db        277 AAAAAEAKKKAVKESSIR------SVQETVLPIKK----------RKTRET--------V 312

Qy        361 KVEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYF 418
              :|| |  : |     |  | |  :    || ||          ::: |: :
Db        313 SIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSPKGR---------- 354
```

```
Qy        419 SSKYNKEALSPPRRK 433
                  :  | |||::: 
Db        355 ----SSSASSPPKKE 365
```

RESULT 13
O42403
ID   O42403   PRELIMINARY;   PRT;   344 AA.
AC   O42403;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Attachment region binding protein (Fragment).
GN   Name=ARBP;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97415642; PubMed=9271441;
RA   Weitzel J.M., Buhrmester H., Straetling W.H.;
RT   "Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG
RT   binding protein MeCP2.";
RL   Mol. Cell. Biol. 17:5656-5666(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20188769; PubMed=10723722;
RA   Reichwald K., Theisen J., Wiehe T., Weitzel J., Straetling W.H.,
RA   Kioschis P., Poustka A., Rosethal A., Platzer M.;
RT   "Comparative sequence analysis of the MECP2-locus in human and mouse
RT   reveals new transcribed regions.";
RL   Mamm. Genome 11:182-190(2000).
DR   EMBL; Y14166; CAA74577.1; -.
DR   PDB; 1UB1; NMR; A=64-196.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000637; A+T_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   InterPro; IPR000976; Wilms_tumour.
DR   Pfam; PF01429; MBD; 1.
DR   PRINTS; PR00929; ATHOOK.
DR   PRINTS; PR00049; WILMSTUMOUR.
DR   SMART; SM00391; MBD; 1.
FT   NON_TER   344   344
SQ   SEQUENCE   344 AA; 33640 MW; 53DD7BD9C9DF4FE3 CRC64;

  Query Match   6.6%; Score 202.5; DB 2; Length 344;
  Best Local Similarity   34.7%; Pred. No. 0.0002;
  Matches   59; Conservative   17; Mismatches   57; Indels   37; Gaps   5;

```
Qy         16 AAPTVTSSERLVPDPPNDLRKEDVAMELER-----VGEDEEQMMIKRSSECNPLLQEPIA 70
              ||      |||       :   | ||   ||      |  | :   :  :|  |   ||
Db          5 AAAAAGGEERL-----EEQADEGVAGLKERPPKAKKGRKERREDPEAEAEAEPSGAEPAE 59

Qy         71 SAQFGATAGTEC-------------RKSV-------------PCGWERVVKQRLFGKTAG 104
              : :   : ||               |:|:             | || | :|||  |::||
Db         60 AGKADGSGGTAAAPAVPEASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAG 119

Qy        105 RFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSR 154
              ::||| |:|||  |||| |  |  | |:||| | |||||| : ||  ||
Db        120 KYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTV-TGRGSPSR 168
```

RESULT 14
MEC2_RAT
ID   MEC2_RAT   STANDARD;   PRT;   492 AA.
AC   Q00566;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN   Name=Mecp2;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=92298389; PubMed=1606614;
RA   Lewis J.D., Meehan R.R., Henzel W.J., Maurer-Fogy I., Jeppesen P.,
RA   Klein F., Bird A.;
RT   "Purification, sequence, and cellular localization of a novel
RT   chromosomal protein that binds to methylated DNA.";
RL   Cell 69:905-914(1992).
CC   -!- FUNCTION: Chromosomal protein that binds to methylated DNA. It can
CC   bind specifically to a single methyl-CpG pair. It is not
CC   influenced by sequences flanking the methyl-CpGs. Mediates
CC   transcriptional repression through interaction with histone
CC   deacetylase and the corepressor SIN3A (By similarity).
CC   -!- SUBUNIT: Interacts with FNBP3 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear. Colocalized with methyl-CpG in the
CC   genome.
CC   -!- TISSUE SPECIFICITY: Present in all adult somatic tissues tested.
CC   -!- SIMILARITY: Contains 2 A.T hook DNA-binding repeats.
CC   -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M94064; AAA41584.1; -.
DR   PIR; A41907; A41907.
DR   HSSP; P51608; 1QK9.
DR   RGD; 3075; Mecp2.
DR   InterPro; IPR000637; A+T_hook.
DR   InterPro; IPR001739; Methyl-CpG_bind.
DR   Pfam; PF01429; MBD; 1.
DR   SMART; SM00384; AT_hook; 2.
DR   SMART; SM00391; MBD; 1.
DR   PROSITE; PS50982; MBD; 1.
KW   Direct protein sequencing; DNA-binding; Nuclear protein; Repeat;
KW   Repressor; Transcription regulation.
FT   DOMAIN   90   162   MBD.
FT   DNA_BIND   185   197   A.T hook 1 (By similarity).
FT   DNA_BIND   265   277   A.T hook 2 (By similarity).
FT   DOMAIN   366   372   His-rich.
FT   DOMAIN   379   403   Pro-rich.
SQ   SEQUENCE   492 AA; 53047 MW; A67E705C68BA2D38 CRC64;

  Query Match   6.6%; Score 201; DB 1; Length 492;
  Best Local Similarity   22.9%; Pred. No. 0.00039;
  Matches 108; Conservative   57; Mismatches 156; Indels 150; Gaps   18;

```
Qy         34 LRKE-----------DVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG------- 75
              ||||           :  :: ::| :|:::    :      |        |: |
Db          8 LRKEKSEDQDLQGLKEKPLKFKKVKKDKKEDKEGKHEPLQPSAHHSAEPAEAGKAETSES 67

Qy         76 ---------ATAGTECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISP 113
                      |:| : |:|:             | || | :|||  |::||::||| |:|
Db         68 SGSAPAVPEASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINP 127

Qy        114 QGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNN 173
              ||  |||| |  |  | |:||| | |||||| : ||  ||
Db        128 QGKAFRSKVELIAYFEKVGDTSLDPNDFDFTV-TGRGSPSR------------------- 167

Qy        174 SNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKV-RKPK 232
                | :  ||     | |       ||      :     :   : |  ::|  |
Db        168 -----REQKPPKK-----PKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEKSP 217
```

```
Qy        233 GKVTILKGIPIKKTKKGCRKSCSGFVQSDS-----KRESVCNKADAESEPVAQKSQLDRT 287
              ||  :|  :| : :  |  |   |  : :     ||     ||:|: : : :|
Db        218 GK--LLVKMPFQASPGG--KGEGGGATTSAQVMVIKRPGRKRKAEADPQAIPKKR----- 268

Qy        288 VCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEK 347
                    |  ::  : |      || |:       | |:|   | |             |
Db        269 -----GRKPGSVVAAAAAEAKKKAVKESSIR------SVQETVLPIKK----------RK 307

Qy        348 YEDTFLESSEEIGTKVEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIP 405
                :|         :|| |  : |   |  | |  :   || ||        ::: |
Db        308 TRET--------VSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSP 351

Qy        406 RTQIERRKTSLYFSSKYNKEALSPPRR-----------KAFKKWTPPRSP 444
              : :             :  | |||::           ||     ||  |
Db        352 KGR-------------SSSASSPPKKEHHHHHHHAESPKAPMPLLPPPPP 388
```

RESULT 15
MEC2_MOUSE
ID MEC2_MOUSE STANDARD; PRT; 484 AA.
AC Q9Z2D6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
GN Name=Mecp2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
RT CpG binding proteins.";
RL Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99299240; PubMed=10369871; DOI=10.1093/hmg/8.7.1253;
RA Coy J.F., Sedlacek Z., Baechner D., Delius H., Poustka A.;
RT "A complex pattern of evolutionary conservation and alternative
RT polyadenylation within the long 3'-untranslated region of the methyl-
RT CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene
RT expression.";
RL Hum. Mol. Genet. 8:1253-1262(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Reichwald K., Thiessen J., Wiehe T., Kioschis P., Straetling W.H.,
RA Rosenthal A., Platzer M.;
RT "Comparative analysis of the methyl CpG binding protein 2 locus in man
RT and mouse reveals new untranslated sequences.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP INTERACTION WITH FNBP3.
RX MEDLINE=97315177; PubMed=9171351; DOI=10.1093/emboj/16.9.2376;
RA Bedford M.T., Chan D.C., Leder P.;
RT "FBP WW domains and the Abl SH3 domain bind to a specific class of
RT proline-rich ligands.";
RL EMBO J. 16:2376-2383(1997).
CC -!- FUNCTION: Chromosomal protein that binds to methylated DNA. It can
CC bind specifically to a single methyl-CpG pair. It is not
CC influenced by sequences flanking the methyl-CpGs. Mediates
CC transcriptional repression through interaction with histone
CC deacetylase and the corepressor SIN3A (By similarity).
CC -!- SUBUNIT: Interacts with FNBP3.
CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalized with methyl-CpG in the
CC genome.
CC -!- SIMILARITY: Contains 2 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AF072251; AAC68880.1; -.
DR EMBL; AJ132922; CAB46495.1; -.
DR EMBL; AF121351; -; NOT_ANNOTATED_CDS.
DR EMBL; AF158181; AAF33024.1; -.
DR EMBL; BC027153; AAH27153.1; -.
DR HSSP; P51608; 1QK9.
DR MGD; MGI:99918; Mecp2.
DR GO; GO:0005634; C:nucleus; IDA.
DR InterPro; IPR000637; A+T_hook.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00384; AT_hook; 2.
DR SMART; SM00391; MBD; 1.
DR PROSITE; PS50982; MBD; 1.
KW DNA-binding; Nuclear protein; Repeat; Repressor;
KW Transcription regulation.
FT DOMAIN 90 162 MBD.
FT DNA_BIND 185 197 A.T hook 1 (By similarity).
FT DNA_BIND 265 277 A.T hook 2 (By similarity).
FT DOMAIN 366 372 His-rich.
FT DOMAIN 379 403 Pro-rich.
SQ SEQUENCE 484 AA; 52307 MW; 62FD228F0118A49F CRC64;

Query Match 6.5%; Score 198.5; DB 1; Length 484;
Best Local Similarity 24.1%; Pred. No. 0.00054;
Matches 91; Conservative 50; Mismatches 154; Indels 83; Gaps 11;

```
Qy         76 ATAGTECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKS 122
              |:|  : |:|:              | || | :|||  |::||::||| |:|||  ||||
Db         77 ASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKV 136

Qy        123 SLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRS 182
              |  |  | |:||| | |||||| : ||  || :           : :|  :    | |
Db        137 ELIAYFEKVGDTSLDPNDFDFTV-TGRGSPSRREQ------KPPKKPKSPKAPGTGRGRG 189

Qy        183 KCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGK------VT 236
              :       | |   :     |       :|:: |   |       | ||       |
Db        190 R--------PKGSGTGRPKAAASEGVQVKRVLEKSPGKLVVKMPFQASPGGKGEGGGATT 241

Qy        237 ILKGIPIKKTKKGCRKSCSGFVQSDSKRE-----SVCNKADAESEPVAQKS----QLDRT 287
                : : ||  : | ::      |:  |:      ||   | ||::  | |      :  |
```

```
Db        242 SAQVMVIK--RPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVHET 299

Qy        288 VCISDAGACGETLSVTSEE-------NSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAK 340
              |          ||:|:  :|       ::| :|  : | :  :   : | |    :  ||
Db        300 VLPIKKRKTRETVSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRKSKESSPKGRSSSAS 359

Qy        341 DSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILK-----------RGSEMDNNCSP 388
                  | :                      | |::  |               ::  ||
Db        360 SPPKKEHHH------------------HHHHSESTKAPMPLLPSPPPPEPESSEDPISP 400

Qy        389 TRKDFTGEKIFQEDTIPR 406
                    | :|: :||
Db        401 PEPQDLSSSICKEEKMPR 418
```

Search completed: August 22, 2005, 10:09:03
Job time : 182 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 22, 2005, 10:02:08 ; Search time 42 Seconds
(without alignments)
1328.708 Million cell updates/sec

Title: US-10-629-951-2
Perfect score: 3055
Sequence: 1 MGTTGLESLSLGDRGAAPTV.....HKLNKYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	209	6.8	476	2	S57963	methyl CpG binding
2	201	6.6	492	2	A41907	methyl-CpG-binding
3	145	4.7	1301	2	S51323	SAC3 protein - yea
4	145	4.7	3418	1	G02334	breast cancer tumo
5	141.5	4.6	782	2	S27833	rhoptry-associated
6	138.5	4.5	500	2	S55785	nucleolar protein
7	137	4.5	560	2	S53382	protein YKR029c ho
8	136.5	4.5	598	2	B40713	cylicin I - human
9	136.5	4.5	1359	2	T34036	hypothetical prote
10	135	4.4	822	2	T41622	probable ABC trans
11	132.5	4.3	286	2	A72682	probable A/G-speci
12	132	4.3	1819	2	A71928	cag island protein
13	131.5	4.3	997	2	T43523	cut17 protein - fi
14	130	4.3	1927	2	G64585	cag pathogenicity
15	128.5	4.2	946	2	A96748	hypothetical prote
16	128	4.2	853	2	T51505	hypothetical prote
17	126	4.1	561	2	H86442	unknown protein [i
18	124.5	4.1	491	2	T50346	hypothetical prote
19	124	4.1	650	2	T33350	hypothetical prote
20	124	4.1	1702	2	T14050	protein kinase (EC
21	123	4.0	697	2	C97120	topoisomerase I [i
22	122.5	4.0	6669	2	S55024	nebulin, skeletal
23	122	4.0	651	2	C86333	hypothetical prote
24	122	4.0	3924	2	S37431	ankyrin 2, neurona
25	120.5	3.9	911	2	S51441	hypothetical prote
26	120.5	3.9	1147	2	JN0599	DNA-binding protei
27	120.5	3.9	1233	2	S56271	hypothetical prote
28	120	3.9	1131	2	A49393	activator 1 large
29	120	3.9	1440	2	T33813	hypothetical prote
30	120	3.9	1658	2	S55101	hypothetical prote
31	120	3.9	2253	2	T30336	nuclear/mitotic ap
32	119.5	3.9	1740	2	T43773	hypothetical prote
33	119.5	3.9	3122	2	T17202	DNA-directed DNA p
34	119	3.9	646	2	F71620	hypothetical prote
35	119	3.9	797	2	H84919	hypothetical prote
36	119	3.9	991	2	H86168	hypothetical prote
37	119	3.9	1002	2	T30546	major surface glyc
38	118.5	3.9	891	2	B84614	hypothetical prote
39	118.5	3.9	1040	2	E71412	hypothetical prote
40	118	3.9	533	2	B84590	hypothetical prote
41	118	3.9	734	2	B42680	nucleolus-cytoplas
42	118	3.9	1200	2	A46194	neurofilament prot
43	118	3.9	1377	2	T51447	transcription regu
44	118	3.9	3329	2	T42205	breast cancer susc
45	118	3.9	3329	2	T30904	breast cancer tumo

ALIGNMENTS

RESULT 1
S57963
methyl CpG binding protein 2 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Nov-1999
C;Accession: S57963
R;d'Esposito, M.; Quaderi, N.A.; Ciccodicola, A.; Bruni, P.; Esposito, T.; D'Urso, M.; B
submitted to the EMBL Data Library, July 1995
A;Description: Physical mapping and expression analysis of an X-linked gene encoding a m
A;Reference number: S57963
A;Accession: S57963
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-476 <DES>
A;Cross-references: EMBL:X89430; NID:g899295; PIDN:CAA61599.1; PID:g899296

Query Match 6.8%; Score 209; DB 2; Length 476;
Best Local Similarity 24.0%; Pred. No. 6.5e-06;
Matches 104; Conservative 53; Mismatches 149; Indels 128; Gaps 17;

```
Qy         37 EDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG----------ATAGTE 81
              :|  :: ::| :|:::    :     |        |: |          |:|  :
Db         13 KDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSARLCEASASPK 72

Qy         82 CRKSV----------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYL 128
              |:|:           | || | :|||  |::||::||| |:||| ||||  |  |
Db         73 QRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYF 132

Qy        129 HKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDV 188
              | |:||| | |||||| : ||  ||                  | :   ||
Db        133 EKVGDTSLDPNDFDFTV-TGRGSPSR----------------------REQKPPKK-- 165

Qy        189 FMPPSSSSELQESRGL---SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIKK 245
              | |       ||       :|       |||     | : |  ||  :|  :|  :
Db        166 ---PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF-Q 217

Qy        246 TKKGCRKSCSGFVQSDS----KRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLS 301
              | | :   |   |      ||      ||:|: : : :|              |  ::
Db        218 TSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKR----------GRKPGSVVA 267

Qy        302 VTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGTK 361
              : |      || |:       | |:|   | |              |  :|
Db        268 AAAAEAKKKAVKESSIR------SVQETVLPIKK----------RKTRET--------VS 303

Qy        362 VEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFS 419
              :|| |  : |    |  | |  :    || ||         ::: |: :
Db        304 IEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSPKGR---------- 344

Qy        420 SKYNKEALSPPRRK 433
              :  | |||:::
Db        345 ---SSSASSPPKKE 355
```

```
RESULT 2
A41907
methyl-CpG-binding protein 2 - rat
N;Alternate names: chromosomal protein MeCP2
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: A41907; S41461
R;Lewis, J.D.; Meehan, R.R.; Henzel, W.J.; Maurer-Fogy, I.; Jeppesen, P.; Klein, F.; Bir
Cell 69, 905-914, 1992
A;Title: Purification, sequence, and cellular localization of a novel chromosomal protei
A;Reference number: A41907; MUID:92298389; PMID:1606614
A;Accession: A41907
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-492 <LEW>
A;Cross-references: UNIPROT:Q00566; GB:M94064; NID:g205361; PIDN:AAA41584.1; PID:g205362
R;Nan, X.; Meehan, R.R.; Bird, A.
Nucleic Acids Res. 21, 4886-4892, 1993
A;Title: Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2.
A;Reference number: S41461; MUID:94232813; PMID:8177735
A;Contents: annotation; methyl CpG-binding domain
C;Keywords: chromosomal protein; DNA binding
F;78-162/Domain: methyl-CpG-binding #status experimental <MCG>

  Query Match             6.6%;  Score 201;  DB 2;  Length 492;
  Best Local Similarity   22.9%;  Pred. No. 2.2e-05;
  Matches  108;  Conservative   57;  Mismatches  156;  Indels  150;  Gaps   18;

Qy         34 LRKE-----------DVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG------- 75
                ||||           :  :: ::| :|:::    :     |         |: |
Db          8 LRKEKSEDQDLQGLKEKPLKFKKVKKDKKEDKEGKHEPLQPSAHHSAEPAEAGKAETSES 67

Qy         76 ---------ATAGTECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISP 113
                     |:|  : |:|:              | || | :||| |::||::||| |:|
Db         68 SGSAPAVPEASASPKQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINP 127

Qy        114 QGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNN 173
              ||  |||| |  |  | |:||| | |||||| : ||  ||
Db        128 QGKAFRSKVELIAYFEKVGDTSLDPNDFDFTV-TGRGSPSR------------------- 167

Qy        174 SNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKV-RKPK 232
                | :   ||     | |        ||      :     :    : |  ::|  |
Db        168 -----REQKPPKK-----PKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEKSP 217

Qy        233 GKVTILKGIPIKKTKKGCRKSCSGFVQSDS-----KRESVCNKADAESEPVAQKSQLDRT 287
              ||  :|  :| : :   |  |    |  : :     ||      ||:|: : : :|
Db        218 GK--LLVKMPFQASPGG--KGEGGGATTSAQVMVIKRPGRKRKAEADPQAIPKKR----- 268

Qy        288 VCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEK 347
                    |  ::  : |       || |:       | |:|   | |              |
Db        269 -----GRKPGSVVAAAAAEAKKKAVKESSIR------SVQETVLPIKK----------RK 307

Qy        348 YEDTFLESEEIGTKVEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIP 405
                :|          :|| |  : |    |  | |  :    || ||         ::: |
Db        308 TRET--------VSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSP 351

Qy        406 RTQIERRKTSLYFSSKYNKEALSPPRR------------KAFKKWTPPRSP 444
              : :              :  | |||::            ||     ||  |
Db        352 KGR-------------SSSASSPPKKEHHHHHHHAESPKAPMPLLPPPPP 388


RESULT 3
S51323
SAC3 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YD8358.13; protein YDR159w
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
C;Accession: S51323; S57983; S71744
R;Bauer, A.; Koelling, R.
submitted to the EMBL Data Library, January 1995
A;Description: The SAC3 gene codes for a nuclear protein required for normal mitosis.
A;Reference number: S51323
A;Accession: S51323
A;Molecule type: DNA
A;Residues: 1-1301 <BAU>
A;Cross-references: UNIPROT:P46674; EMBL:Z47805; NID:g634085; PID:g634086
R;Murphy, L.; Richards, C.; Harris, D.
submitted to the EMBL Data Library, July 1995
A;Reference number: S57971
A;Accession: S57983
A;Molecule type: DNA
A;Residues: 1-1301 <MUR>
A;Cross-references: EMBL:Z50046; NID:g899393; PID:g899406; MIPS:YDR159w
A;Experimental source: strain AB972
R;Bauer, A.; Koelling, R.
Yeast 12, 965-975, 1996
A;Title: Characterization of the SAC3 gene of Saccharomyces cerevisiae.
A;Reference number: S71744; MUID:97027306; PMID:8873450
A;Accession: S71744
A;Molecule type: DNA
A;Residues: 1-1301 <BAW>
A;Cross-references: EMBL:Z47805; NID:g634085; PIDN:CAA87767.1; PID:g634086
C;Genetics:
A;Gene: SGD:SAC3; LEP1
A;Cross-references: SGD:S0002566; MIPS:YDR159w
A;Map position: 4R
C;Function:
A;Description: potential regulator of leucine permease gene expression
C;Keywords: nucleus; transmembrane protein
F;999-1015/Domain: transmembrane #status predicted <TMM>

  Query Match             4.7%;  Score 145;  DB 2;  Length 1301;
  Best Local Similarity   21.9%;  Pred. No. 0.32;
  Matches  116;  Conservative   61;  Mismatches  185;  Indels  168;  Gaps   25;

Qy        123 SLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDC----------------------S 159
              :| :| ||  ||     : :  | | :|  |: ||                       |
Db        510 TLQHYSHKLSETQPLKKTY-LTCLERRLQKTTYKGLINGGEDNLASSVYVKDPKKDRIPS 568

Qy        160 MAALTSHLQNQSNNSNWNLRTRSKCK-------KDVFMPPSSSSELQESRGLSNFTSTHL 212
              :|  :  ::|  || |  |    | |       | |   |:     |||: |   : :|
Db        569 IADQSFLMENFQNNYNEKLNQNSSVKPQINTSPKRVATRPNHFPFSQESKQLPQISQSHT 628

Qy        213 LLKEDEGVDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKAD 272
              |          |     :  | ::  |  ||    |     | ||   | : |     :
Db        629 L--------STNPLLTPQVHGDLSEQKQQQIKTVTDG----GSPFVFDQSAQNSTVEASK 676

Qy        273 AESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGI 332
              |              :  :  ||  | ||  ||:  : ||:|: :         |:||
Db        677 A------------HMISTTSNGAYDEKLS--SEQEEMRKKEEQRI-------EEEKT--- 712

Qy        333 INKFCSAKDSEHNEKYEDTFLESEEIGTKV--EVVERKEHLHTDILKRGSEMDNNCSPTR 390
                  |  : |     |  : :|:|    :  |||           |:||      |     |
Db        713 -----QLKKKQEN---ADKQVITEQIANDLVKEVVNSSV---ISIVKREFSEAN----YR 757

Qy        391 KDFTGEKIFQEDTIPR----TQIERRKTSLYFSSKYNKEALSPPRRKAFKKW-------- 438
              |||         ||: |       :  |   :|  |:   :  |  ::| |:||
Db        758 KDFI-------DTMTRELYDAFLHERLYLIYMDSRAELKRNSTLKKKFFEKWQASYSQAK 810

Qy        439 ------TPPRSPFNLVQE-----------TLFHDPWKLLIATIFLNRTSGKMAI--PV-- 477
                    |    ||               ||  |:|  : : |:  :| |   |  ||
Db        811 KNRILEEKKREEIKLVSHQLGVPGFKKSTCLFRTPYKGNVNSSFMLSSSDKNLIFSPVND 870

Qy        478 --------------LWKFLE-----------KYPSAEVARTADWRDVSELLKPLGLYDLR 512
                           ||: ||           |:|           |  | |  |: :
Db        871 EFNKFATHLTKISKLWRPLEMQSIYYDNLTKKFP------------SNSLTPANLF-IY 916

Qy        513 AKTIVKFSDEYLTKQWKYPIELHGIGKYGND--SYRIFCVNEWKQVHPED 560
              ||     |: ::  ::          |: |:  | || |:::  :  | |
Db        917 AKDWTSLSNRWILSKFNLQ-TAQDSKKFSNNIISSRIICIDD--EYEPSD 963


RESULT 4
G02334
breast cancer tumor suppressor BRCA2 - human
```

N;Alternate names: breast cancer susceptibility protein BRCA2
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C;Accession: G02334; S68501
R;Tavtigian, S.V.; Rommens, J.M.; Couch, F.J.; Neuhausen, S.; Bell, R.; Berry, S.; Bogde
er, M.; Snyder, S.; Stringfellow, M.; Stroup, C.; Swedlund, B.; Teng, D.; Thomas, A.; Tr
submitted to the EMBL Data Library, December 1995
A;Reference number: H01078
A;Accession: G02334
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3418 <TAV>
A;Cross-references: UNIPROT:P51587; EMBL:U43746; NID:g1161383; PIDN:AAB07223.1; PID:g116
R;Wooster, R.; Bignell, G.; Lancaster, J.; Swift, S.; Seal, S.; Mangion, J.; Collins, N.
ith, A.; Connor, F.; Arason, A.; Gudmundsson, J.; Ficenec, D.; Kelsell, D.; Ford, D.; To
; Narod, S.; Lenoir, G.; Egilsson, V.; Barkadottir, R.B.; Easton, D.F.; Bentley, D.R.
Nature 378, 789-792, 1995
A;Authors: Futreal, P.A.; Ashworth, A.; Stratton, M.R.
A;Title: Identification of the breast cancer susceptibility gene BRCA2.
A;Reference number: S68501; MUID:96112016; PMID:8524414
A;Accession: S68501
A;Molecule type: mRNA
A;Residues: 282-371,'N',373-598,'S',600-1108,'EQ',1111-1119,'D',1121-2321,'V',2323-2386,
C;Genetics:
A;Gene: GDB:BRCA2
A;Cross-references: GDB:387848; OMIM:600185
A;Map position: 13q12.3-13q12.3
C;Superfamily: DNA recombination repair protein, BRCA2 type
C;Keywords: polymorphism; tumor suppressor

Query Match 4.7%; Score 145; DB 1; Length 3418;
Best Local Similarity 20.9%; Pred. No. 1.2;
Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

```
Qy        36 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVK 95
             | |::: :::: |      | | :      | ||::  || :  |   | :      : |
Db       972 KSDISLNIDKIPEKNNDYMNKWAG-----LLGPISNHSFGGSFRTASNKEIKLSEHNIKK 1026

Qy        96 QRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSL-KPEDFDFTVLSKRGIKSR 154
              ::| |     |:    |  |   :   : | |  : :  | ||:  : || :
Db      1027 SKMFFK-----DIEEQYPTSL---ACVEIVNTLALDNQKKLSKPQSIN-TVSAHLQSSVV 1077

Qy       155 YKDCSMAALTSHL--QNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHL 212
               ||  : :|  :    |  ||| ||             ||  :|: |         || |
Db      1078 VSDCKNSHITPQMLFSKQDFNSNHNL-----------TPSQKAEITE-------LSTIL 1118

Qy       213 LLKEDEGVDDVNFRKVRK-----------PKGKVTILK--------------------- 239
               :|      | : ||             |: ::||||
Db      1119 ----EESGSQFEFTQFRKPSYILQKSTFEVPENQMTILKTTSEECRDADLHVIMNAPSIG 1174

Qy       240 ----------GIPIKK-----TKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQL 284
                       : ||:      |  | || ||::  :       |:       |  ::|
Db      1175 QVDSSKQFEGTVEIKRKFAGLLKNDCNKSASGYLTDE-------NEVGFRGFYSAHGTKL 1227

Qy       285 D-RTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSE 343
             :  |  :  |      :  |||  |    :  :|  |: | :     ::: |   |
Db      1228 NVSTEALQKAVKLFSDIENISEETS---AEVHPISLSSSKCHD----SVVSMF---KIEN 1277

Qy       344 HNEKYEDTFLESEE-------------IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
             ||:|   |  |                  || || :       |:  || :| ::|   :
Db      1278 HNDK---TVSEKNNKCQLILQNNIEMTTGTFVEEI-------TENYKRNTENEDNKYTAA 1327

Qy       389 TRK----DFTGEKIFQEDTIPRTQIERRKTSLYFSSKYN 423
             :|     :| |    : ||:    | : :| | |: ::|
Db      1328 SRNSHNLEFDGSDSSKNDTV---CIHKDETDLLFTDQHN 1363
```

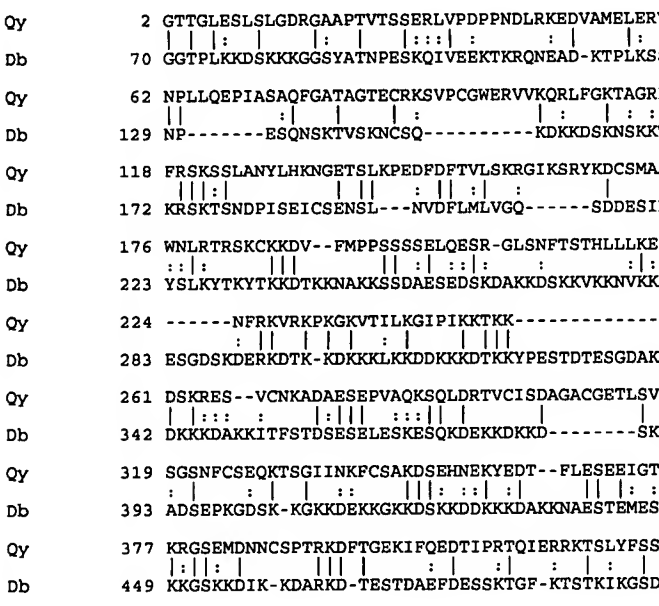
RESULT 5
S27833
rhoptry-associated protein 1 precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: protective antigen
C;Species: Plasmodium falciparum
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: A45514; S27833
R;Ridley, R.G.; Takacs, B.; Lahm, H.W.; Delves, C.J.; Goman, M.; Certa, U.; Matile, H.;
Mol. Biochem. Parasitol. 41, 125-134, 1990
A;Title: Characterisation and sequence of a protective rhoptry antigen from Plasmodium f
A;Reference number: A45514; MUID:90348711; PMID:2200961
A;Accession: A45514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-782 <RI2>
A;Cross-references: UNIPROT:Q26007; GB:M32853; NID:g160656; PID:g160657
C;Superfamily: Plasmodium falciparum rhoptry-associated protein 1

Query Match 4.6%; Score 141.5; DB 2; Length 782;
Best Local Similarity 21.1%; Pred. No. 0.27;
Matches 110; Conservative 71; Mismatches 176; Indels 165; Gaps 25;

```
Qy       123 SLANYLHKNGET----SLKPEDFDF------TVLSKRGIKSRYKDCSMAALTSHLQNQS- 171
             ::|: :: ||:     ::   ||:|      | ::|:   : |:|    :  | |:|:|
Db        18 NVADGINVNGDNNYGKTIINNDFNFDDYNYWTPINKKEFLNSYED--EFSSESFLENKSS 75

Qy       172 -NNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRK 230
              :: | ||    |   |         ||:    |       |   :|:||:  ||: |
Db        76 VDDGNINLTDTSTSNK--------SSKKGHGRSRVRSASAAAILEEDDSKDDMEF----- 122

Qy       231 PKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCI 290
             |   :::|      |:    || |    | :|  | |                      :
Db       123 -KASPSVVKTSTPSGTQTSGLKSSS---PSSTKSSSPSN--------------------V 158

Qy       291 SDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYED 350
               |   ||:    | | |  |  :|| |          :||:      | |     ::
Db       159 KSASPHGES---NSSEESTTKSSKRSAS----------VAGIV-----GADEEAPPAPKN 200

Qy       351 TFLESEEI-GTKVEVVERKEHLHT-----DILKRGS---------EMDNNCSPTRKDFTG 395
             |    ||:  | | :   |  |:      :|||              | | |    ::|
Db       201 TLTPLEELYPTNVNLFNYKYSLNNMEENINILKNEGDLVAQKEEFEYDENMEKAKQDKKK 260

Qy       396 --EKI----------FQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRS 443
               |||          | |:      |:: |  :  ||  ::|  |:|      |||
Db       261 ALEKIGKQSDEEPFMFSENKFLENQVKERNVAGSFSRFFSK--LNP-----FKK------ 307

Qy       444 PFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSA--------------- 488
               ::::|   :  |   : |  | |  :  :  :    ::||
Db       308 -DEVIEKT---EVSKKTFSGIGFNLTDKEAKVLGVGATYQEYPETMLYNCPNNSNLFDTI 363

Qy       489 -------------EVARTADWRDVSELLKPLGLYDLRAK-TIVKFSDEYLTKQWKYPIEL 534
                         |   :  :    | || :|: ||     |  ||
Db       364 ESLQGRIIDIKKRESMISTTFEQQKECLKNMGVLDLELNDTQCKFGT------------- 410

Qy       535 HGIGKYGNDSYRIF-CVNEWKQVHP--------EDHKLNKYH 567
               || :|    |::    |:  : ||          : :|| | |
Db       411 -CIGSFGEHHLRLYEFENDLFKFHPNIDYLTLADGYKLQKNH 451
```

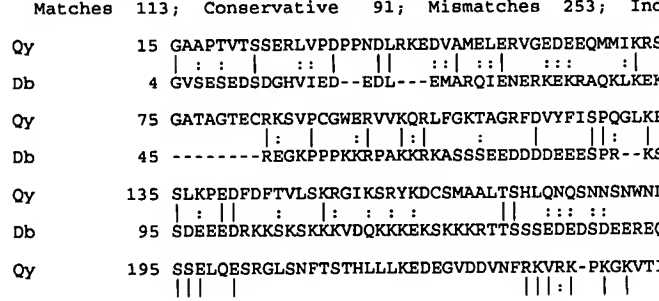
RESULT 6
S55785
nucleolar protein gar2 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: probable RNA-binding protein RBD18
C;Species: Schizosaccharomyces pombe
C;Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 16-Aug-2004
C;Accession: S55785; S68087; T37634; S52318
R;Gulli, M.P.; Girard, J.P.; Zabetakis, D.; Lapeyre, B.; Melese, T.; Caizergues-Ferrer,
Nucleic Acids Res. 23, 1912-1918, 1995
A;Title: gar2 is a nucleolar protein from Schizosaccharomyces pombe required for 18S rR
A;Reference number: S55785; MUID:95319932; PMID:7596817
A;Accession: S55785
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-500 <GUL>
A;Cross-references: UNIPROT:P41891; EMBL:Z48166; NID:g663261; PIDN:CAA88179.1; PID:g663
R;VanHoy, R.W.; Wise, J.A.
Curr. Genet. 29, 307-315, 1996
A;Title: Molecular analysis of a novel Schizosaccharomyces pombe gene containing two RN
A;Reference number: S68083; MUID:96171513; PMID:8598051

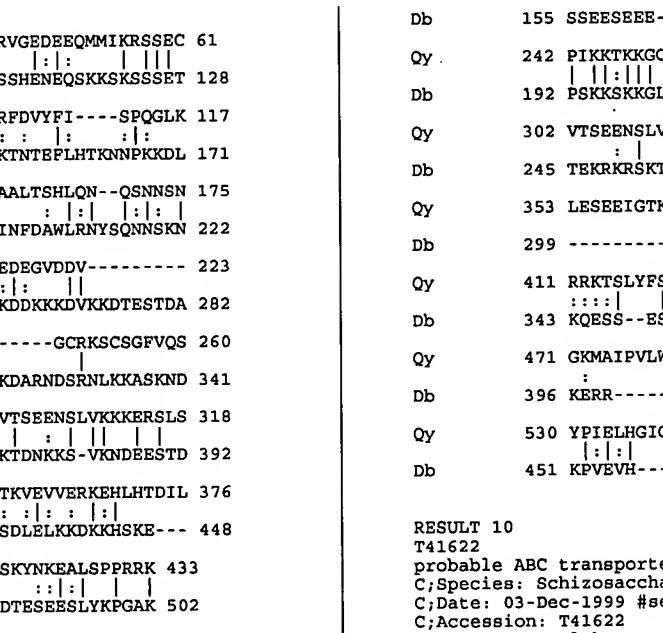
```
A;Accession: S68087
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 374-407 <VAN>
R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21733
A;Accession: T37634
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-338,'S',340-500 <BRO>
A;Cross-references: EMBL:Z99091; PIDN:CAB11772.1; GSPDB:GN00066; SPDB:SPAC13F5.09
A;Experimental source: strain 972h-; cosmid c13F5
C;Genetics:
A;Gene: gar2; SPDB:SPAC13F5.09
A;Map position: 1
C;Superfamily: ribonucleoprotein repeat homology
F;264-331/Domain: ribonucleoprotein repeat homology <RRM1>
F;367-433/Domain: ribonucleoprotein repeat homology <RRM2>

  Query Match             4.5%;  Score 138.5;  DB 2;  Length 500;
  Best Local Similarity   23.7%;  Pred. No. 0.23;
  Matches   74;  Conservative   50;  Mismatches  147;  Indels   41;  Gaps   11;

Qy        185 KKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIK 244
Db         19 KKGAIEKPSKSKKITKEAAKE--------IAKQSSKTDVSPKKSKKEAKRASSPE--PSK 68

Qy        245 KTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLSVTS 304
Db         69 KSVKKQKKSKK---KEESSSESESSSSSESESSSSESESSSS---ESESSSSESSSSES 122

Qy        305 EENSLVK---KKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGTK 361
Db        123 EEEVIVKTEEKKESSSESSSSSSESEEEEEAVV-KIEEKKESSSDSSSESSSSESESESSS 181

Qy        362 VEVVERKEHLH-TDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
Db        182 SESEEEEEVVEKTEEKKEGSSESSSDSESSSDSSSE---SGDSDSSSDSESESS----SE 234

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
Db        235 DEKKRKAEPASEERPAKITKPSQDSN---ET----------CTVFVGRLSWNVDDQWLGQ 281

Qy        481 FLEKYPSAEVAR 492
Db        282 EFEEYGTIVGAR 293


RESULT 7
S53382
protein YKR029c homolog YJL105w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein J0819
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S53382; S56883; S57363
R;Rasmussen, S.W.
submitted to the EMBL Data Library, February 1995
A;Description: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and
A;Reference number: S53376
A;Accession: S53382
A;Molecule type: DNA
A;Residues: 1-560 <RAS>
A;Cross-references: UNIPROT:P42948; EMBL:X85021; NID:g728698; PID:g728705
R;Rasmussen, S.W.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56876
A;Accession: S56883
A;Molecule type: DNA
A;Residues: 1-560 <RAW>
A;Cross-references: EMBL:Z49380; NID:g1008285; PID:g1008286; MIPS:YJL105w
R;Rasmussen, S.W.
Yeast 11, 873-883, 1995
A;Title: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and CSD3 ge
A;Reference number: S57357; MUID:96090136; PMID:7483851
A;Accession: S57363
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-560 <RAF>
A;Cross-references: EMBL:X85021; NID:g728698; PIDN:CAA59389.1; PID:g728705
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
C;Genetics:
A;Cross-references: SGD:S0003641
A;Map position: 10L

  Query Match             4.5%;  Score 137;  DB 2;  Length 560;
  Best Local Similarity   18.7%;  Pred. No. 0.33;
  Matches   90;  Conservative   80;  Mismatches  161;  Indels  150;  Gaps   24;

Qy        118 FRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHL--QNQSNNSN 175
Db         82 FQNDRGIFNHHSSSGSSKT-------ASTNKRGIAAAVALATAATIPFPLKKQNQDDNSK 134

Qy        176 WNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKV 235
Db        135 VSV-THNESSKENKITPSMRAE-------DNKPKNGCICGSSDSKDEL------------ 174

Qy        236 TILKGIPIKKTKKGCRKSCSGFVQSDS-KRESVCNKADAESEPVAQKSQLD--------- 285
Db        175 ----FIQCNKCKTWQHKLCYAFKKSDPIKRDFVCKRCDSDTK--VQVNQVKPMIFPRKMG 228

Qy        286 -------RTVCISDAGACGE-------------------TLSVTSEENSLVKK-KERSLS 318
Db        229 DERLFQFSSIVTTSASNTNQHQQSVNNIEEQPKKRQLHYTAPTTENSNSIRKKLRQEKLV 288

Qy        319 SGSNFCSEQKTSGIINKFCSAKD--------SEHNEKYEDTFLESEEIGTKVEVVERKEH 370
Db        289 VSSHFLKP-----LLNEVSSSNDTEFKAITISEYKDKYVKMFIDN-------------H 329

Qy        371 LHTDIL-------KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQ-IERRKTSLYFSSKY 422
Db        330 YDDDWVVCSNWESSRSADIEVRKSSNERDF---GVFAADSCVKGELIQEYLGKIDFQKNY 386

Qy        423 NKEALSPPRRKAFKKWTPPRSPFNLVQET----LFHDPWKLLIATIFLNRTSGKMAIPVL 478
Db        387 Q--------------TDPNNDYRLMGTTKPKVLFHPHWPLYIDS---RETGG------L 422

Qy        479 WKFLEK--YPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKF-----SDEYLTKQWKYP 531
Db        423 TRYIRRSCEPNVELVTV---RPLDE--KPRGDNDCRVKFVLRAIRDIRKGEEISVEWQWD 477

Qy        532 I 532
Db        478 L 478


RESULT 8
B40713
cylicin I - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 09-Jul-2004
C;Accession: B40713; S35920
R;Hess, H.; Heid, H.; Franke, W.W.
J. Cell Biol. 122, 1043-1052, 1993
A;Title: Molecular characterization of mammalian cylicin, a basic protein of the sperm he
A;Reference number: A40713; MUID:93359502; PMID:8354692
A;Accession: B40713
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-598 <HES>
A;Cross-references: UNIPROT:P35663; GB:Z22780; NID:g396104; PIDN:CAA80457.1; PID:g396105
C;Keywords: cytoskeleton

  Query Match             4.5%;  Score 136.5;  DB 2;  Length 598;
  Best Local Similarity   23.3%;  Pred. No. 0.39;
  Matches  111;  Conservative   72;  Mismatches  205;  Indels   89;  Gaps   21;
```

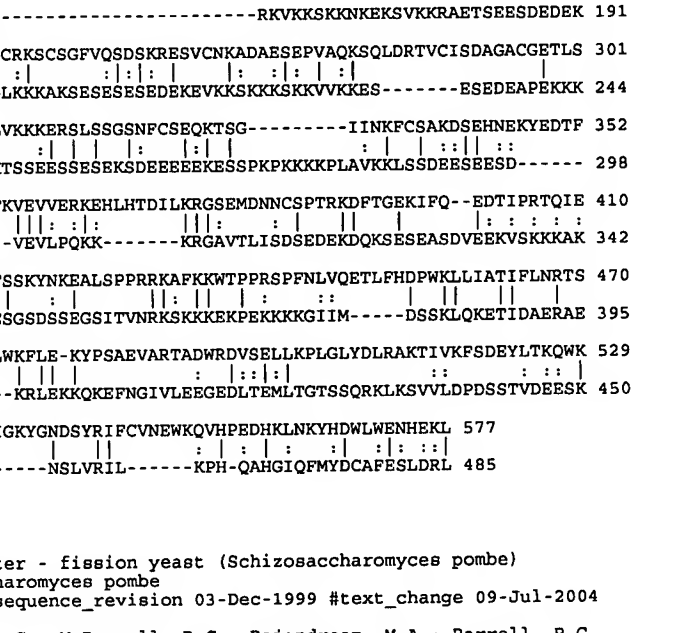


RESULT 9
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T34036
R;Fulton, R.; Wohldmann, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid B0041.
A;Reference number: Z21466
A;Accession: T34036
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1359 <FUL>
A;Cross-references: UNIPROT:Q9U7E0; EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:
A;Experimental source: strain Bristol N2; clone B0041
C;Genetics:
A;Gene: CESP:B0041.7
A;Map position: 1
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 4.5%; Score 136.5; DB 2; Length 1359;
Best Local Similarity 19.2%; Pred. No. 1.2;
Matches 113; Conservative 91; Mismatches 253; Indels 131; Gaps 19;

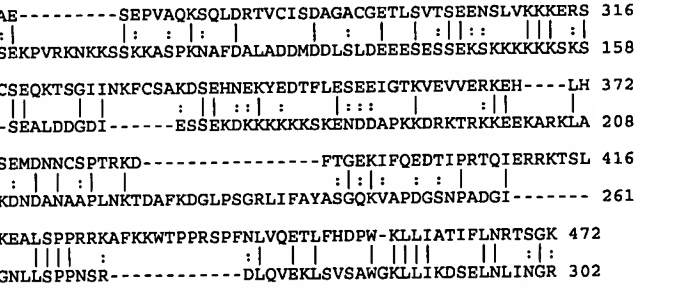






RESULT 10
T41622
probable ABC transporter - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T41622
R;Aert, R.; Volckaert, G.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21735
A;Accession: T41622
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-822 <AER>
A;Cross-references: UNIPROT:Q9USH9; EMBL:AL122011; PIDN:CAB58409.1; GSPDB:GN00068; SPDB:
A;Experimental source: strain 972h-; cosmid c825
C;Genetics:
A;Gene: SPDB:SPCC825.01
A;Map position: 3

Query Match 4.4%; Score 135; DB 2; Length 822;
Best Local Similarity 22.4%; Pred. No. 0.75;
Matches 80; Conservative 53; Mismatches 134; Indels 90; Gaps 14;



```
RESULT 11
A72682
probable A/G-specific adenine glycosylase APE0875 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: A72682
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: A72682
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-286 <KAW>
A;Cross-references: UNIPROT:Q9YDP0; DDBJ:AP000060; NID:g5104188; PIDN:BAA79857.1; PID:d1
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0875

  Query Match             4.3%;  Score 132.5;  DB 2;  Length 286;
  Best Local Similarity   29.8%;  Pred. No. 0.26;
  Matches   28;  Conservative   26;  Mismatches   31;  Indels    9;  Gaps    3;

Qy        455 DPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAK 514
              ||| :|:|    | :|: :  : | :|| :||: :   :|   :| ||::|||:   |||
Db         98 DPWAILVAAFLLRKTTARQVVRVYEEFLRRYPNPKALASAREDEVRELIRPLGIEHQRAK 157

Qy        515 TIVKFSDEYLTKQW--KYPI------ELHGIGKY 540
               ::: : :::  ::  : |       || |:| |
Db        158 HLIELA-KHIEARYGGRIPCSKEKLKELPGVGDY 190


RESULT 12
A71928
cag island protein - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C;Accession: A71928
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: A71928
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1819 <ARN>
A;Cross-references: UNIPROT:Q9ZLV0; GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD0604
A;Experimental source: strain J99
C;Genetics:
A;Gene: orf13/14

  Query Match             4.3%;  Score 132;  DB 2;  Length 1819;
  Best Local Similarity   18.2%;  Pred. No. 3.4;
  Matches   87;  Conservative   83;  Mismatches  183;  Indels  124;  Gaps   19;

Qy          6 LESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGED---------------- 49
              ||: :|     | |    :| | | | ||:|: :| |  ||  |
Db        942 LENQALDCLKNAKTEAEKKRCVKDLPKDLQKKVLAKESVRVYLDCVSKAKNEAERKECEK 1001

Qy         50 ----EEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGR 105
                   | : ::: : |     :: :: |:       || | :     ::::| :
Db       1002 LLTPEARKLLEEAKESVKAYKDCVSRAR-NEKEKQECEKLLTPEARKLLEQEVKKSVKAY 1060

Qy        106 FDVY-------------FISPQGLKFRSKSSLANYLHKNGETSLK--------PEDFDF 143
               |             ::|:  |     :|     || :|  :         |:|
Db       1061 LDCVSRARNEKEKQECEKLLTPEARKLLENQALD--CLKNAKTEAEKKRCVKDLPKDLQK 1118

Qy        144 TVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRG 203
              ||:|  :|: | ||      |  :|:        : : :|:|   : | :   |:||:
Db       1119 KVLAKESVKA-YLDC-----VSRARNE--------KEKKECEK--LLTPEARKLLEESK- 1161

Qy        204 LSNFTSTHL--LLKEDEGVDDVNFRKVRKPKGKVTILKGIPIKKTKKGC---------- 250
                    :|  : |     :     |:  |: :  : :     |  | |
Db       1162 --KSVKAYLDCVSKAKNEAEKKECEKLLTPEARKLLEEAKESVKAYKDCVSRARNEKEKQ 1219

Qy        251 ------------------RKSCSGFVQSDSKRESVCNKADAES--EPVA------QKSQL 284
                               :||   ::   |:  :   | : |    | |       |: |
Db       1220 ECEKLLTPEARKLLEQEVKKSVKAYLDCVSRARNEKEKQECEKLLTPEARKFLEKQRQQK 1279

Qy        285 DRTV--CI-----SDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFC 337
              |: :  |:     :|  | : |   |:|  |              :  | :   :::
Db       1280 DKAIKDCLKNADPNDRAAIMKCLDGLSDEEKL------------KYLQEAREKAVLDCLK 1327

Qy        338 SAKDSEHNEKYEDTFLE-SEEIGTKVEVVERKEHLHTDILKRGSE-MDNNCSPTRKD 392
              :|:  |   | :: : :  :||  |    :: :   |: | : || :||   || ::
Db       1328 TARTDEEKRKCQNLYSDLIQEIQNKKAQNKQNQLSKTERLHQASECLDNLDDPTDQE 1384


RESULT 13
T43523
cut17 protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T43523; T41649; T41700
R;Morishita, J.; Matsusaka, T.; Yanagida, M.
submitted to the EMBL Data Library, August 1999
A;Description: Fission yeast cut17 is required for chromosome segregation.
A;Reference number: Z22536
A;Accession: T43523
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-997 <MOR>
A;Cross-references: UNIPROT:O14064; EMBL:AB031034; PIDN:BAA83415.1
R;Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z22007
A;Accession: T41649
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-997 <HAR>
A;Cross-references: EMBL:AL031323; PIDN:CAA20434.1; GSPDB:GN00068; SPDB:SPCC962.02c
A;Experimental source: strain 972h-; cosmid c962
R;Wedler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z22010
A;Accession: T41700
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 932-997 <WED>
A;Cross-references: EMBL:AL121859; PIDN:CAB58376.1; GSPDB:GN00068; SPDB:SPCP31B10.10c
C;Genetics:
A;Gene: cut17; SPCC962.02c; SPDB:SPCP31B10.10c
A;Map position: 3L
A;Introns: 43/3

  Query Match             4.3%;  Score 131.5;  DB 2;  Length 997;
  Best Local Similarity   20.1%;  Pred. No. 1.6;
  Matches  129;  Conservative  119;  Mismatches  241;  Indels  153;  Gaps   33;

Qy         20 VTSSERLVPDPPNDLRKEDVAMEL---ERVGEDEEQMMIKRS----SECNPLLQEPIASA 72
              ::||    |     |    |: ||:|    ::| :::| : :: |     |    |:  :|  |:
Db        416 ISSSVSSVGKEQNHTEKQ-VAIETPEQQKVEKEDEHLNLQGSFIEESTKQPISSKPSTSS 474

Qy         73 QFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNG 132
                  | |  | |      :::::     ::           |  :  :  |  ||   :|
Db        475 PDMTDAATGGRVSSSSFRDKILQTNFSPRST------IDSFSNISKKRNSEEAN--DEND 526

Qy        133 ETSLK---PE---DFDFTVLSKRGIKSRYKDC---------SMAAL----TSHLQNQSNN 173
              ||:||   ||     |   : ||  : |  :|          |  |:       |:|:|
Db        527 ETNLKIPIPEKKRKFQEVLQSKNILVSSTEDSHEPVKVTEDSQTAIHVSKFEDLENKSME 586
```

```
Qy       174 SNWNLRTRSKCKKD----VFMPPSSSSELQES-------RGLSNFTSTHLLLKEDEGVDD 222
Db       587 SEQSLQLLSESENDDKPLIDLIPLLAIKRKDNLVSGVLEKGKSTSTSK---TKFDTSIVD 643

Qy       223 VNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQ--------------SDSKRESV 267
Db       644 F----IEKPKTEIS--EVLPEEKRKAICDESQTVRVSIDRGVTKTRDVSSPVSDEKSENV 697

Qy       268 CNKADAES------------EPVAQKSQLDRTVCISDA--GACGETLSVTSEENSLVKK 312
Db       698 -NHEEANSGHTVMNVHSSLDPQPIVQPNELESGSYLKDLPDRNVGNSEKVTFQEDDINSP 756

Qy       313 KERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLH 372
Db       757 KLQSKNNQTVEAVNTETSDKLQEKEANHELENIEKIEEKLTEVDKVSLSDAFPDQEIKNS 816

Qy       373 TDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERR----KTSLYFSSKY-NKEAL 427
Db       817 RTSVQNGTRSVSKNTPEKE----TKVDKIDNVSKKDVETSPGSCETSSAFAKTYAEKEVT 872

Qy       428 S---PPRRKAFKK--WTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFL 482
Db       873 SINLPSVRKPLDESYYDHSISPFDPLCQSSFLAP---------QTPVKSKHALPLV---- 919

Qy       483 EKYPSAEVARTADWR--DVSELLK-----PLGLYDLRAKTIVKFSDEYLTKQW-KYPIEL 534
Db       920 -------EANAPPWEPIDFSSLLESPVPNPVEPNKLSEKEL-----DMTVEQWIKF---- 963

Qy       535 HGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEK 576
Db       964 ----MYAK------CAKEFEEACEE--KI----EWLLEEGKR 989
```

RESULT 14
G64585
cag pathogenicity island protein cag7 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: G64585
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D. Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: G64585
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1927 <TOM>
A;Cross-references: UNIPROT:O25262; GB:AE000567; GB:AE000511; NID:g2313641; PIDN:AAD0759

Query Match 4.3%; Score 130; DB 2; Length 1927;
Best Local Similarity 20.5%; Pred. No. 5;
Matches 89; Conservative 61; Mismatches 147; Indels 138; Gaps 21;

```
Qy        17 APTVTSSERLVPDPPNDLRKEDVAME-----LERVGEDEEQMMIKRSSECNPLL----QE 67
Db      1137 AKTEADKKRCVKDLPKDLQKKVLAKESVKAYLDCVSRARNE---KEKKECEKLLTPEAKK 1193

Qy        68 PIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGR---------FDVY--------- 109
Db      1194 LLEEAKESLKAYKDCLSQARNEEERRACEKLLTPEARKLLEQEVKKSIKAYLDCVSRARN 1253

Qy       110 ---------FISPQGLKFRSKSSLANYLHKNG-----ETSLK--PEDFDFTVLSKRGIKS 153
Db      1254 EKEKKECEKLLTPEARKFLAKQVL-NCLEKAGNEEERKACLKNLPKDLQENILAKESLKA 1312

Qy       154 RYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLL 213
Db      1313 -YKDC-----LSQARNEEE--------RRACEK--LLTPEARKLLEQ------------- 1343

Qy       214 LKEDEGVDDVNFRKVRKP-KGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKAD 272
Db      1344 -------------EVKKSVKAYLDCVSRARNEKEKKECEKLLT----------------- 1373

Qy       273 AESEPVAQK------SQLDRTV--CI-----SDAGACGETLSVTSEENSLVKKKERSLSS 319
Db      1374 ----PEARKFLAKELQQKDKAIKDCLKNADPNDRAAIMKCLDGLSDEEKL---------- 1419

Qy       320 GSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLE-SEEIGTKVEVVERKEHLHTDILKR 378
Db      1420 --KYLQEAREKAVADCLAMAKTDEEKRKCQNLYSDLIQEIQNKRTQNKQNQLSKTERLHQ 1477

Qy       379 GSE-MDNNCSPTRKD 392
Db      1478 ASECLDNLDDPTDQE 1492
```

RESULT 15
A96748
hypothetical protein T10D10.13 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: A96748
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A96748
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-946 <STO>
A;Cross-references: UNIPROT:Q9C9D8; GB:AE005173; NID:g6730761; PIDN:AAF27150.1; GSPDB:GN
C;Genetics:
A;Gene: T10D10.13
A;Map position: 1

Query Match 4.2%; Score 128.5; DB 2; Length 946;
Best Local Similarity 21.0%; Pred. No. 2.4;
Matches 94; Conservative 56; Mismatches 177; Indels 121; Gaps 16;

```
Qy        32 NDLRKEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCG-- 89
Db       370 NELEKLFAEHQLRVPGDQSSSSRRGKSSENQVAHKEPSHS-----IAATEKRLSLGGGSA 424

Qy        90 -WERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSK 148
Db       425 DFSKLMTPLVGDKDKGDALRRNLSDLSLTDDSKGKFYEKYMKKRDAKLREE--------- 475

Qy       149 RGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFT 208
Db       476 WSLKKGEKETKLKSMQEALEQSRTEMKAKLSASSSERKDLL-----SSTRQRAEKFRSFN 530

Qy       209 STHLLLK----EDEGVDDVNFR----------------------------------- 226
Db       531 SRSSMKKYQSEEDEDISEQKPRAKDKAASGQQSVGSISSRSSQARKLQPNRNMSSSITPR 590

Qy       227 ---KVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQ 283
Db       591 SAASVPKPSGKVS--------NTSSGRRRS------DKSLAQSVPNFSELIKENTKPSSL 636

Qy       284 LDRTVCISDAGACGETLSVTSEENSLVKKKE--RSLSSG-------SNFCSEQKTSGIIN 334
Db       637 AVKTTMRSQVKSSGRTKNI--KEDTLLQRPRSLRKSSSGNIDFTELSTLCSDDM---MVS 691

Qy       335 KFCSAKDSE--HNEKYEDTFLESEEI-GTKVEVVERKEHLHTDILKRGSEM--------- 382
Db       692 LRVDSDISETLRNEEYDEPEAEPEEVLENAVREEEEVEELETLVFEDGNPMLSEAYEKVD 751
```

```
Qy        383 ---DNNCSPTRKDFTGEKIFQEDTIPRT 407
                  : |||           |   |:| |
Db        752 HSGEENCS-----------FLPATVPTT 768
```

Search completed: August 22, 2005, 10:06:00
Job time : 46 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 22, 2005, 10:02:08 ; Search time 170 Seconds
(without alignments)
1319.535 Million cell updates/sec

Title: US-10-629-951-2
Perfect score: 3055
Sequence: 1 MGTTGLESLSLGDRGAAPTV.....HKLNKYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3055	100.0	580	2	AAW74473	Aaw74473 Human MED
2	3055	100.0	580	7	ADD89906	Add89906 Human 5-m
3	3055	100.0	580	8	ADK13954	Adk13954 Human met
4	2958.5	96.8	565	3	AAY44504	Aay44504 Human del
5	2194.5	71.8	439	2	AAY76548	Aay76548 Human ova
6	1639	53.6	307	7	ADD89915	Add89915 Human 5-m
7	1230	40.3	257	2	AAW88701	Aaw88701 Secreted
8	1230	40.3	257	4	ABB50468	Abb50468 Human sec
9	1230	40.3	257	6	ABO44725	Abo44725 Novel hum
10	1230	40.3	257	7	ABO26205	Abo26205 Human pro
11	1106	36.2	202	7	ADD89916	Add89916 Human 5-m
12	880.5	28.8	416	7	ADD89908	Add89908 Chicken 5
13	816	26.7	147	7	ADD89917	Add89917 Human 5-m
14	357	11.7	68	5	AAE22568	Aae22568 Human MBD
15	254	8.3	50	4	ABB51147	Abb51147 Human sec
16	254	8.3	50	6	ABO45404	Abo45404 Novel hum
17	254	8.3	50	7	ABO26884	Abo26884 Protein a
18	220.5	7.2	467	8	ADK13968	Adk13968 Frog meth
19	220.5	7.2	467	8	ADK13967	Adk13967 Frog meth
20	213.5	7.0	486	8	ADK13972	Adk13972 Human met
21	209	6.8	476	7	ADJ68460	Adj68460 Human hea
22	209	6.8	476	8	ADK13973	Adk13973 Human met
23	208.5	6.8	477	8	ADK13961	Adk13961 Human met
24	208.5	6.8	486	8	ADK13966	Adk13966 Human met
25	208.5	6.8	486	8	ADK13971	Adk13971 Human met
26	208.5	6.8	486	8	ADK13957	Adk13957 Human met
27	208.5	6.8	486	8	ADK13960	Adk13960 Human met
28	208.5	6.8	486	8	ADK13953	Adk13953 Human met
29	208.5	6.8	486	8	ADK13963	Adk13963 Human met
30	208.5	6.8	486	8	ADK13970	Adk13970 Human met
31	208.5	6.8	486	8	ADK13958	Adk13958 Human met
32	208.5	6.8	486	8	ABM82465	Abm82465 Tumour-as
33	208.5	6.8	560	6	ABG72561	Abg72561 TAT dMT-M
34	208.5	6.8	561	6	ABG72557	Abg72557 MeCP2-Tat
35	202.5	6.6	345	8	ADK14010	Adk14010 Chicken m
36	201	6.6	492	8	ADK13964	Adk13964 Rat methy
37	198.5	6.5	484	8	ADK13959	Adk13959 Mouse met
38	198.5	6.5	484	8	ADK13956	Adk13956 Human met
39	198.5	6.5	484	8	ADK13969	Adk13969 Mouse met
40	198.5	6.5	484	8	ADK13965	Adk13965 Mouse met
41	194	6.4	219	3	AAG02051	Aag02051 Human sec
42	186	6.1	68	5	AAE22566	Aae22566 Human MeC
43	186	6.1	92	8	ADK13962	Adk13962 Human met
44	159.5	5.2	744	8	ADP49805	Adp49805 Human 193
45	148.5	4.9	326	3	AAG16859	Aag16859 Arabidops

ALIGNMENTS

RESULT 1
AAW74473
ID AAW74473 standard; protein; 580 AA.
XX
AC AAW74473;
XX
DT 19-MAY-1999 (first entry)
XX
DE Human MED1 endonuclease protein sequence.
XX
KW Endonuclease; MED1; human; methyl-CpG binding endonuclease-1;
KW DNA fidelity; DNA manipulation; cancer; fragile X syndrome; therapy;
KW myotonic dystrophy; Huntington's disease; spinocerebellar ataxia;
KW Kennedy's disease; triplet repeat expansion disorder.
XX
OS Homo sapiens.
XX
PN WO9904626-A1.
XX
PD 04-FEB-1999.
XX
PF 28-JUL-1998; 98WO-US015828.
XX
PR 28-JUL-1997; 97US-0053936P.
XX
PA (FOXC-) FOX CHASE CANCER CENT.
XX
PI Bellacosa A;
XX
DR WPI; 1999-142462/12.
DR N-PSDB; AAX22002.
XX
PT New nucleic acid encoding human endonuclease MED1 involved in DNA
PT mismatch repair - used for diagnosing susceptibility to cancer and
PT fragile X syndrome, and therapeutically.
XX
PS Claim 8; Fig 3; 109pp; English.
XX
CC This sequence is the human MED1 endonuclease of the invention. MED1 (for
CC methyl-CpG binding endonuclease-1) is used to screen for specific
CC modulators (potential therapeutic agents particularly mimetics of MED1)
CC and to study interactions involved in maintaining DNA fidelity, for DNA
CC manipulation and to raise antibodies. Susceptibility or predisposition to
CC cancer (particularly colorectal or endometrial, especially hereditary non
CC -polyposis colorectal cancer), or its prognosis, where caused by
CC alterations in the MED1-encoding gene, are identified by sequence
CC comparison, amplification, detecting altered polypeptide, and restriction
CC fragment mapping, hybridisation (particularly to probes specific for a

```
CC   mutant allele). These same methods can also be used to diagnose fragile X
CC   syndrome and other diseases (e.g, myotonic dystrophy, Huntington's
CC   disease, spinocerebellar ataxia and Kennedy's disease) associated with
CC   triplet repeat expansion. The DNA, or its fragments, are used as probes
CC   and primers in the above diagnostic methods, also to isolate homologous
CC   sequences, as sources of antisense sequences and for gene transfer,
CC   particularly to restore drug sensitivity to drug-resistant cancer cells
XX
SQ   Sequence 580 AA;

  Query Match             100.0%;  Score 3055;  DB 2;  Length 580;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-287;
  Matches  580;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              ||||||||||||||||||||||||||||||||||||||||
Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580


RESULT 2
ADD89906
ID   ADD89906 standard; protein; 580 AA.
XX
AC   ADD89906;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human 5-methylcytosine DNA glycosylase.
XX
KW   Human; 5-methylcytosine DNA glycosylase; enzyme; CpG.
XX
OS   Homo sapiens.
XX
PN   WO2003078593-A2.
XX
PD   25-SEP-2003.
XX
PF   14-MAR-2003; 2003WO-US007933.
XX
PR   15-MAR-2002; 2002US-0364689P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Lofton-Day CE,  Day JK;
XX
DR   WPI; 2003-779127/73.
DR   N-PSDB; ADD89905.
XX
PT   Labeling methylated or methylatable CpG sequences, useful e.g. for
PT   diagnostic detection of altered methylation, comprises replacing
PT   methylated cytosine by labeled cytosine.
XX
PS   Claim 11; Page 53-55; 73pp; English.
XX
CC   The present sequence is the protein sequence of human 5-methylcytosine
CC   DNA glycosylase (5-MCDG). The enzyme acts by cleaving glycosylic bonds at
CC   methylated CpG sites of DNA, removing 5-methylcytosine from the DNA
CC   backbone as a free base. Human 5-MCDG can be used in a claimed method for
CC   labelling CpG sequences corresponding to methylated CpG sequences in an
CC   isolated DNA sample. The method comprises: digesting the genomic DNA with
CC   a restriction endonuclease to produce genomic DNA fragments; treating the
CC   genomic DNA fragments with 5-MCDG such that one or more 5-methylcytosine
CC   bases are removed to produce abasic genomic DNA fragments; and treating
CC   these abasic genomic DNA fragments with base excision repair enzymes in
CC   the presence of labelled dCTP such that 5-methylcytosine removed from the
CC   genomic DNA fragments by 5-MCDG is replaced by labelled cytosine in the
CC   one or more corresponding positions of the abasic genomic DNA fragments
CC   to produce labelled genomic DNA fragments, so that specific labelling of
CC   CpG sequences corresponding to methylated CpG sequences is achieved. The
CC   5-MCDG is also used in a claimed method for comparing CpG methylation
CC   status, extent or pattern between or among reference and test genomic DNA
CC   samples, and in a claimed method for labelling potentially-methylatable
CC   CpG sequences in CpG-containing genomic DNA fragments. The methods are
CC   used to identify methylated and/or potentially methylatable CpG
CC   dinucleotides in genomic DNA, including comparison of methylation pattern
CC   between healthy and diseased samples, for diagnosis.
XX
SQ   Sequence 580 AA;

  Query Match             100.0%;  Score 3055;  DB 7;  Length 580;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-287;
  Matches  580;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
```

```
Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              |||||||||||||||||||||||||||||||||||||||||
Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580


RESULT 3
ADK13954
ID   ADK13954 standard; protein; 580 AA.
XX
AC   ADK13954;
XX
DT   03-JUN-2004  (first entry)
XX
DE   Human methyl-CpG-binding protein #9.
XX
KW   Rett syndrome; methyl-CpG-binding protein 2; MECP2;
KW   neurodevelopmental disease; autism; non-syndromic mental retardation;
KW   idiopathic neonatal encephalopathy; idiopathic infantile spasm;
KW   idiopathic cerebral palsy; Angelman syndrome; schizophrenia; human.
XX
OS   Homo sapiens.
XX
PN   US6709817-B1.
XX
PD   23-MAR-2004.
XX
PF   07-SEP-2000; 2000US-00657013.
XX
PR   07-SEP-1999;   99US-0152778P.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Zoghbi HY,  Van Den Veyver IB,  Amir R,  Francke U;
XX
DR   WPI; 2004-256068/24.
XX
PT   Screening human for Rett syndrome comprises detecting mutation in nucleic
PT   acid sequence encoding methyl-CpG-binding protein 2 (MECP2).
XX
PS   Disclosure; SEQ ID NO 56; 125pp; English.
XX
CC   The invention relates to a method of screening a human for Rett syndrome
CC   comprising detecting a mutation in a nucleic acid sequence encoding
CC   methyl-CpG-binding protein 2 (MECP2). The method is useful for screening
CC   a human for Rett syndrome. The method is useful for screening
CC   neurodevelopmental diseases such as Rett syndrome, autism, non-syndromic
CC   mental retardation, idiopathic neonatal encephalopathy, idiopathic
CC   infantile spasms, idiopathic cerebral palsy, Angelman syndrome and
CC   schizophrenia. The present sequence represents the amino acid sequence of
CC   a methyl-CpG-binding protein.
XX
SQ   Sequence 580 AA;

  Query Match             100.0%;  Score 3055;  DB 8;  Length 580;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-287;
  Matches  580;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              |||||||||||||||||||||||||||||||||||||||||
Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580


RESULT 4
AAY44504
ID   AAY44504 standard; protein; 565 AA.
XX
AC   AAY44504;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Human delta228-UV damage endonuclease.
XX
KW   Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW   glutathione-S-transferase signal peptide; uve1+ gene product;
KW   UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW   abasic site; aplatinum diaduct; mismatched nucleotide pairing;
KW   nucleotide alkylation; skin cancer.
XX
OS   Homo sapiens.
XX
PN   WO9963828-A1.
XX
PD   16-DEC-1999.
XX
PF   08-JUN-1999;   99WO-US012910.
XX
PR   08-JUN-1998;   98US-0088521P.
PR   18-MAY-1999;   99US-0134752P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Doetsch PW,  Kaur B,  Avery AM;
XX
DR   WPI; 2000-116417/10.
XX
PT   A new truncated ultraviolet damage endonuclease for treatment of skin
PT   cancers.
XX
PS   Claim 16; Page 60; 133pp; English.
XX
CC   The present sequence is human delta228-UV damage endonuclease. Delta228-
```

```
CC   UVDE is a 288 amino acid deletion of the N-terminal of the uve1+ gene
CC   product. This is expressed in frame with a GST leader sequence to
CC   generate a fusion protein. This provides stable endonuclease fragments
CC   for cleaving a double-stranded DNA molecule that has a distorted
CC   structure resulting from UV radiation damage, a photoproduct, an abasic
CC   site, mismatched nucleotide pairing, aplatinum diaduct, an intercalated
CC   molecule or alkylation of a nucleotide. Uve1p can be used in compositions
CC   for internal or topical application and as a therapeutic agent for skin
CC   cancers
XX
SQ   Sequence 565 AA;

  Query Match             96.8%;  Score 2958.5;  DB 3;  Length 565;
  Best Local Similarity   97.4%;  Pred. No. 1.7e-277;
  Matches  565;  Conservative    0;  Mismatches    0;  Indels   15;  Gaps    1;

Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTV------------ 288

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 ---SEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 345

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 405

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 465

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 525

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              |||||||||||||||||||||||||||||||||||||||
Db        526 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 565


RESULT 5
AAY76548
ID   AAY76548 standard; protein; 439 AA.
XX
AC   AAY76548;
XX
DT   10-APR-2000  (first entry)
XX
DE   Human ovarian tumor EST fragment encoded protein 44.
XX
KW   Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KW   gene therapy; treatment.
XX
OS   Homo sapiens.
XX
PN   DE19817557-A1.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1998;   98DE-01017557.
XX
PR   09-APR-1998;   98DE-01017557.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-591920/51.
DR   N-PSDB; AAZ77469.
XX
PT   New nucleic acid sequences expressed in ovarian, and some other, cancer
PT   tissues, and derived polypeptides, for treatment of ovarian cancer and
PT   identification of therapeutic agents.
XX
PS   Claim 25; Page 261; 310pp; German.
XX
CC   This invention describes novel nucleic acid (cDNA) sequences (A) which
CC   have anticancer activity and are highly expressed in ovarian tumor tissue
CC   (and some also in testis and breast cancer tissue). The products of the
CC   invention can be used for gene therapy. (A) are used (i) for recombinant
CC   expression of polypeptides (B) and (ii) to isolate complete genes. (B)
CC   are used (i) to identify agents suitable for treatment of ovarian cancer;
CC   (ii) directly for treating this form of cancer (including expression from
CC   gene therapy vectors) and (iii) for generation of specific antibodies.
CC   (A) are identified by assembling ESTs (expressed sequence tags) from a
CC   particular tissue type before comparison of expression patterns. This
CC   allows a significantly longer fragment of the gene to be revealed, so
CC   should reduce the number of failures associated with the fact that ESTs
CC   from different libraries may represent different parts of the same
CC   unknown gene, distorting the estimated frequency of occurrence in a
CC   particular tissue. AAY76505-Y76638 represent protein fragments encoded by
CC   the human ovarian tumor cDNA library derived EST fragments represented in
CC   AAZ77450-Z77572
XX
SQ   Sequence 439 AA;

  Query Match             71.8%;  Score 2194.5;  DB 2;  Length 439;
  Best Local Similarity   96.0%;  Pred. No. 1.8e-203;
  Matches  427;  Conservative    3;  Mismatches    6;  Indels    9;  Gaps    3;

Qy         95 KQRLFGKTAGRFD-VYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKS 153
              |  ||  |: :|  : | ||||||||||||||||||||||||||||||||||||||||
Db          1 KSLLF--TSSKFPLISFSSPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKS 58

Qy        154 RYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLL 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         59 RYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLL 118

Qy        214 LKEDEGVDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADA 273
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 LKEDEGVDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADA 178

Qy        274 ESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGII 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 ESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGII 238

Qy        334 NKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDF 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 NKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDF 298

Qy        394 TGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLF 453
              |    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 T------EDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLF 352

Qy        454 HDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRA 513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        353 HDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRA 412

Qy        514 KTIVKFSDEYLTKQWKYPIELHGIG 538
```

```
                    ||||||||||||||||||||||||
Db        413 KTIVKFSDEYLTKQWKYPIELHGIG 437


RESULT 6
ADD89915
ID   ADD89915 standard; protein; 307 AA.
XX
AC   ADD89915;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human 5-methylcytosine DNA glycosylase N-terminal deletion mutant.
XX
KW   Human; 5-methylcytosine DNA glycosylase; enzyme; CpG; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO2003078593-A2.
XX
PD   25-SEP-2003.
XX
PF   14-MAR-2003; 2003WO-US007933.
XX
PR   15-MAR-2002; 2002US-0364689P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Lofton-Day CE,  Day JK;
XX
DR   WPI; 2003-779127/73.
XX
PT   Labeling methylated or methylatable CpG sequences, useful e.g. for
PT   diagnostic detection of altered methylation, comprises replacing
PT   methylated cytosine by labeled cytosine.
XX
PS   Claim 11; Page 70-71; 73pp; English.
XX
CC   The present sequence is the protein sequence of an N-terminal deletion
CC   mutant of human 5-methylcytosine DNA glycosylase (5-MCDG), in which amino
CC   acid residue 1 corresponds to amino acid 274 of the full-length protein
CC   ADD89906. 5-MCDG acts by cleaving glycosylic bonds at methylated CpG
CC   sites of DNA, removing 5-methylcytosine from the DNA backbone as a free
CC   base. The N-terminal deletion mutant shows enhanced deglycosylase
CC   specificity towards CpG dinucleotide sequences. Human 5-MCDG can be used
CC   in a claimed method for labelling CpG sequences corresponding to
CC   methylated CpG sequences in an isolated DNA sample. The method comprises:
CC   digesting the genomic DNA with a restriction endonuclease to produce
CC   genomic DNA fragments; treating the genomic DNA fragments with 5-MCDG
CC   such that one or more 5-methylcytosine bases are removed to produce
CC   abasic genomic DNA fragments; and treating these abasic genomic DNA
CC   fragments with base excision repair enzymes in the presence of labelled
CC   dCTP such that 5-methylcytosine removed from the genomic DNA fragments by
CC   5-MCDG is replaced by labelled cytosine in the one or more corresponding
CC   positions of the abasic genomic DNA fragments to produce labelled genomic
CC   DNA fragments, so that specific labelling of CpG sequences corresponding
CC   to methylated CpG sequences is achieved. The 5-MCDG is also used in a
CC   claimed method for comparing CpG methylation status, extent or pattern
CC   between or among reference and test genomic DNA samples, and in a claimed
CC   method for labelling potentially-methylatable CpG sequences in CpG-
CC   containing genomic DNA fragments. The methods are used to identify
CC   methylated and/or potentially methylatable CpG dinucleotides in genomic
CC   DNA, including comparison of methylation pattern between healthy and
CC   diseased samples, for diagnosis.
XX
SQ   Sequence 307 AA;

  Query Match           53.6%;  Score 1639;  DB 7;  Length 307;
  Best Local Similarity  100.0%;  Pred. No. 8.4e-150;
  Matches  307;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        274 ESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGII 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db          1 ESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGII 60

Qy        334 NKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDF 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         61 NKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDF 120

Qy        394 TGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLF 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        121 TGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLF 180

Qy        454 HDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRA 513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        181 HDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRA 240

Qy        514 KTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWEN 573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        241 KTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWEN 300

Qy        574 HEKLSLS 580
              |||||||
Db        301 HEKLSLS 307


RESULT 7
AAW88701
ID   AAW88701 standard; protein; 257 AA.
XX
AC   AAW88701;
XX
DT   01-MAR-1999  (first entry)
XX
DE   Secreted protein encoded by gene 168 clone HCFNF11.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9854963-A2.
XX
PD   10-DEC-1998.
XX
PF   04-JUN-1998;   98WO-US011422.
XX
PR   06-JUN-1997;   97US-0048875P.
PR   06-JUN-1997;   97US-0048876P.
PR   06-JUN-1997;   97US-0048877P.
PR   06-JUN-1997;   97US-0048878P.
PR   06-JUN-1997;   97US-0048880P.
PR   06-JUN-1997;   97US-0048881P.
PR   06-JUN-1997;   97US-0048882P.
PR   06-JUN-1997;   97US-0048883P.
PR   06-JUN-1997;   97US-0048884P.
PR   06-JUN-1997;   97US-0048885P.
PR   06-JUN-1997;   97US-0048892P.
PR   06-JUN-1997;   97US-0048893P.
PR   06-JUN-1997;   97US-0048894P.
PR   06-JUN-1997;   97US-0048895P.
PR   06-JUN-1997;   97US-0048896P.
PR   06-JUN-1997;   97US-0048897P.
PR   06-JUN-1997;   97US-0048898P.
PR   06-JUN-1997;   97US-0048899P.
PR   06-JUN-1997;   97US-0048900P.
PR   06-JUN-1997;   97US-0048901P.
PR   06-JUN-1997;   97US-0048915P.
PR   06-JUN-1997;   97US-0048916P.
PR   06-JUN-1997;   97US-0048917P.
```

PR 06-JUN-1997; 97US-0048949P.
PR 06-JUN-1997; 97US-0048962P.
PR 06-JUN-1997; 97US-0048963P.
PR 06-JUN-1997; 97US-0048964P.
PR 06-JUN-1997; 97US-0048970P.
PR 06-JUN-1997; 97US-0048971P.
PR 06-JUN-1997; 97US-0048972P.
PR 06-JUN-1997; 97US-0048974P.
PR 06-JUN-1997; 97US-0049019P.
PR 06-JUN-1997; 97US-0049020P.
PR 06-JUN-1997; 97US-0049373P.
PR 06-JUN-1997; 97US-0049374P.
PR 06-JUN-1997; 97US-0049375P.
PR 05-SEP-1997; 97US-0057584P.
PR 05-SEP-1997; 97US-0057627P.
PR 05-SEP-1997; 97US-0057628P.
PR 05-SEP-1997; 97US-0057629P.
PR 05-SEP-1997; 97US-0057634P.
PR 05-SEP-1997; 97US-0057635P.
PR 05-SEP-1997; 97US-0057642P.
PR 05-SEP-1997; 97US-0057643P.
PR 05-SEP-1997; 97US-0057644P.
PR 05-SEP-1997; 97US-0057645P.
PR 05-SEP-1997; 97US-0057646P.
PR 05-SEP-1997; 97US-0057647P.
PR 05-SEP-1997; 97US-0057648P.
PR 05-SEP-1997; 97US-0057649P.
PR 05-SEP-1997; 97US-0057650P.
PR 05-SEP-1997; 97US-0057651P.
PR 05-SEP-1997; 97US-0057654P.
PR 05-SEP-1997; 97US-0057661P.
PR 05-SEP-1997; 97US-0057662P.
PR 05-SEP-1997; 97US-0057666P.
PR 05-SEP-1997; 97US-0057667P.
PR 05-SEP-1997; 97US-0057668P.
PR 05-SEP-1997; 97US-0057760P.
PR 05-SEP-1997; 97US-0057761P.
PR 05-SEP-1997; 97US-0057762P.
PR 05-SEP-1997; 97US-0057763P.
PR 05-SEP-1997; 97US-0057764P.
PR 05-SEP-1997; 97US-0057765P.
PR 05-SEP-1997; 97US-0057769P.
PR 05-SEP-1997; 97US-0057770P.
PR 05-SEP-1997; 97US-0057771P.
PR 05-SEP-1997; 97US-0057774P.
PR 05-SEP-1997; 97US-0057775P.
PR 05-SEP-1997; 97US-0057776P.
PR 05-SEP-1997; 97US-0057777P.
PR 05-SEP-1997; 97US-0057778P.
PR 18-DEC-1997; 97US-0070923P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Young P, Greene JM, Ferrie AM, Ruben SM, Rosen CA, Hu J;
PI Olsen HS, Ebner R, Brewer LA, Moore PA, Shi Y, Florence C;
PI Florence K, Lafleur DW, Ni J, Fan P, Wei Y, Fischer CL, Soppet DR;
PI Li Y, Zeng Z, Kyaw H, Yu G, Feng P, Dillon PJ, Endress GA;
PI Carter KC;
XX
DR WPI; 1999-059865/05.
DR N-PSDB; AAV84578.
XX
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders.
XX
PS Claim 11; Page 564-565; 772pp; English.
XX
CC The invention relates to nucleic acid sequences (AAV84411 to AAV84633)
CC encoding human secreted proteins (AAW88534 to AAW88756). The secreted
CC protein gene sequences are deposited with the ATCC under deposit numbers
CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC cells comprising recombinant vectors containing the nucleic acid
CC sequences are used for the recombinant production of the secreted
CC proteins. The polynucleotide and amino acid sequences are useful for are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung
CC or thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners. The
CC present sequence represents human secreted protein (see descriptor line
CC for gene number and clone identification)
XX
SQ Sequence 257 AA;

Query Match 40.3%; Score 1230; DB 2; Length 257;
Best Local Similarity 93.1%; Pred. No. 3.3e-110;
Matches 241; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

```
Qy        160 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 60

Qy        220 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVA 279
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVXSDSKRESVCNKADAESEPVA 120

Qy        280 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 180

Qy        340 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIF 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT----- 235

Qy        400 QEDTIPRTQIERRKTSLYF 418
              ||||||    :::    |
Db        236 -EDTIPRNTDRKKENKPVF 253
```

RESULT 8
ABB50468
ID ABB50468 standard; protein; 257 AA.
XX
AC ABB50468;
XX
DT 07-FEB-2002 (first entry)
XX
DE Human secreted protein encoded by gene 168 SEQ ID NO:416.
XX
KW Human; secreted protein; immunomodulatory; antisclerotic; anti-HIV;
KW dermatological; immunosuppressive; antiinflammatory; immunostimulant;
KW cytostatic; cardiant; vascular; anti-angiogenic; ophthalmological;
KW neuroprotective; nootropic; anticonvulsant; antialzheimers; vulnerary;
KW antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;
KW multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;
KW human immunodeficiency virus; hyperproliferative disorder; wound healing;
KW Gaucher's disease; cardiovascular disease; Scimitar syndrome; chemotaxis;
KW Chaga's cardiomyopathy; coronary arteriosclerosis; angiogenic disorder;
KW corneal graft neovascularisation; diabetic retinopathy; regeneration;
KW neurological disorder; Huntington's chorea; Alzheimer's disease;
KW Parkinson's disease; infectious disease; chromosome 3.
XX
OS Homo sapiens.
XX

PN WO200162891-A2.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-US005614.
XX
PR 24-FEB-2000; 2000US-0184836P.
PR 29-MAR-2000; 2000US-0193170P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Ebner R, Lafleur DW, Moore PA, Olsen HS, Rosen CA;
PI Ruben SM, Soppet DR, Young PE, Shi Y, Florence KA, Wei Y;
PI Florence C, Hu J, Li Y, Kyaw H, Fischer CL, Ferrie AM, Fan P;
PI Feng P, Endress GA, Dillon PJ, Carter KC, Brewer LA, Yu G, Zeng Z;
PI Greene JM;
XX
DR WPI; 2001-625724/72.
DR N-PSDB; ABA83361.
XX
PT Nucleic acids encoding 207 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating, e.g. cancers, Parkinson's disease
PT and diabetic retinopathy.
XX
PS Claim 11; Page 1171-1172; 1533pp; English.
XX
CC ABB50301 to ABB51287 and ABA83194 to ABA83441 represent human secreted
CC proteins (I) and polynucleotide (II) sequences. (I) and (II) have various
CC activities based on the tissues and cells the genes are expressed in.
CC Example of these activities include: immunomodulatory; antisclerotic;
CC dermatological; immunosuppressive; antiinflammatory; immunostimulant;
CC anti-HIV; cytostatic; cardiant; anti-angiogenic; ophthalmological;
CC neuroprotective; nootropic; anticonvulsant; antialzheimers; vascular;
CC antiparkinsonian; antimicrobial; and vulnerary. (I) and (II) can be used
CC in gene therapy and vaccine production. (I) and (II) can be used in the
CC prevention, diagnosis and treatment of immune disorders (e.g. multiple
CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus
CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,
CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic
CC disorders (e.g. corneal graft neovascularisation and diabetic
CC retinopathy), neurological disorders (e.g. Huntington's chorea,
CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
CC for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to
CC ABA83193 and ABB50300 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 257 AA;

Query Match 40.3%; Score 1230; DB 4; Length 257;
Best Local Similarity 93.1%; Pred. No. 3.3e-110;
Matches 241; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

```
Qy      160 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 219
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 60

Qy      220 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVA 279
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db       61 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVXSDSKRESVCNKADAESEPVA 120

Qy      280 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 339
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 180

Qy      340 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIF 399
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT----- 235

Qy      400 QEDTIPRTQIERRKTSLYF 418
              ||||||    :::    |
Db      236 -EDTIPRNTDRKKENKPVF 253
```

RESULT 9
ABO44725
ID ABO44725 standard; protein; 257 AA.
XX
AC ABO44725;
XX
DT 02-OCT-2003 (first entry)
XX
DE Novel human secreted protein #168.
XX
KW Human; gene therapy; autoimmune disorder; multiple sclerosis; cancer;
KW systemic lupus erythematosus; haematopoietic cell disorder; allergy;
KW agammaglobulinaemia; ataxia telangiectasia; blood coagulation disorder;
KW afibrinogenaemia; thrombocytopenia; graft-versus-host disease; arthritis;
KW inflammatory condition; ischaemia-reperfusion injury; infectious disease;
KW hyperproliferative disorder; purpura; viral infection; regeneration;
KW bacterial infection; ulcer; Alzheimer's disease.
XX
OS Homo sapiens.
XX
PN US2003065160-A1.
XX
PD 03-APR-2003.
XX
PF 07-DEC-2001; 2001US-00004860.
XX
PR 06-JUN-1997; 97US-0048875P.
PR 06-JUN-1997; 97US-0048876P.
PR 06-JUN-1997; 97US-0048877P.
PR 06-JUN-1997; 97US-0048878P.
PR 06-JUN-1997; 97US-0048880P.
PR 06-JUN-1997; 97US-0048881P.
PR 06-JUN-1997; 97US-0048882P.
PR 06-JUN-1997; 97US-0048883P.
PR 06-JUN-1997; 97US-0048884P.
PR 06-JUN-1997; 97US-0048885P.
PR 06-JUN-1997; 97US-0048892P.
PR 06-JUN-1997; 97US-0048893P.
PR 06-JUN-1997; 97US-0048894P.
PR 06-JUN-1997; 97US-0048895P.
PR 06-JUN-1997; 97US-0048896P.
PR 06-JUN-1997; 97US-0048897P.
PR 06-JUN-1997; 97US-0048898P.
PR 06-JUN-1997; 97US-0048899P.
PR 06-JUN-1997; 97US-0048900P.
PR 06-JUN-1997; 97US-0048901P.
PR 06-JUN-1997; 97US-0048915P.
PR 06-JUN-1997; 97US-0048916P.
PR 06-JUN-1997; 97US-0048917P.
PR 06-JUN-1997; 97US-0048949P.
PR 06-JUN-1997; 97US-0048962P.
PR 06-JUN-1997; 97US-0048963P.
PR 06-JUN-1997; 97US-0048964P.
PR 06-JUN-1997; 97US-0048970P.
PR 06-JUN-1997; 97US-0048971P.
PR 06-JUN-1997; 97US-0048972P.
PR 06-JUN-1997; 97US-0048974P.
PR 06-JUN-1997; 97US-0049019P.
PR 06-JUN-1997; 97US-0049020P.
PR 06-JUN-1997; 97US-0049373P.
PR 06-JUN-1997; 97US-0049374P.
PR 06-JUN-1997; 97US-0049375P.
PR 05-SEP-1997; 97US-0057584P.
PR 05-SEP-1997; 97US-0057627P.
PR 05-SEP-1997; 97US-0057628P.
PR 05-SEP-1997; 97US-0057629P.
PR 05-SEP-1997; 97US-0057634P.
PR 05-SEP-1997; 97US-0057635P.
PR 05-SEP-1997; 97US-0057642P.
PR 05-SEP-1997; 97US-0057643P.
PR 05-SEP-1997; 97US-0057644P.
PR 05-SEP-1997; 97US-0057645P.

```
PR   05-SEP-1997;   97US-0057646P.
PR   05-SEP-1997;   97US-0057647P.
PR   05-SEP-1997;   97US-0057648P.
PR   05-SEP-1997;   97US-0057649P.
PR   05-SEP-1997;   97US-0057650P.
PR   05-SEP-1997;   97US-0057651P.
PR   05-SEP-1997;   97US-0057654P.
PR   05-SEP-1997;   97US-0057661P.
PR   05-SEP-1997;   97US-0057662P.
PR   05-SEP-1997;   97US-0057666P.
PR   05-SEP-1997;   97US-0057667P.
PR   05-SEP-1997;   97US-0057668P.
PR   05-SEP-1997;   97US-0057760P.
PR   05-SEP-1997;   97US-0057761P.
PR   05-SEP-1997;   97US-0057762P.
PR   05-SEP-1997;   97US-0057763P.
PR   05-SEP-1997;   97US-0057764P.
PR   05-SEP-1997;   97US-0057765P.
PR   05-SEP-1997;   97US-0057769P.
PR   05-SEP-1997;   97US-0057770P.
PR   05-SEP-1997;   97US-0057771P.
PR   05-SEP-1997;   97US-0057774P.
PR   05-SEP-1997;   97US-0057775P.
PR   05-SEP-1997;   97US-0057776P.
PR   05-SEP-1997;   97US-0057777P.
PR   05-SEP-1997;   97US-0057778P.
PR   18-DEC-1997;   97US-0070923P.
PR   04-JUN-1998;   98WO-US011422.
PR   15-JUL-1998;   98US-0092921P.
PR   30-JUL-1998;   98US-0094657P.
PR   04-DEC-1998;   98US-00205258.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Young P,  Greene JM,  Ferrie AM,  Ruben SM,  Rosen CA,  Hu J;
PI   Olsen HS,  Ebner R,  Brewer LA,  Moore PA,  Shi Y,  Florence C;
PI   Florence K,  Lafleur DW,  Ni J,  Fan P,  Wei Y,  Fischer CL,  Soppet DR;
PI   Li Y,  Zeng Z,  Kyaw H,  Yu G,  Feng P,  Dillon PJ,  Endress GA;
PI   Carter KC;
XX
DR   WPI; 2003-540804/51.
DR   N-PSDB; ACH04862.
XX
PT   New isolated protein, useful for preparing a composition for diagnosing
PT   or treating cancer, inflammatory, immune or infectious diseases.
XX
PS   Disclosure; SEQ ID NO 416; 172pp; English.
XX
CC   The invention relates to an isolated HEMAE80 protein. The protein is
CC   useful for preparing a composition for diagnosing or treating autoimmune
CC   disorders e.g. multiple sclerosis and systemic lupus erythematosus;
CC   haematopoietic cell disorders e.g. agammaglobulinaemia and ataxia
CC   telangiectasia; blood coagulation disorders e.g. afibrinogenaemia and
CC   thrombocytopenia; allergy; graft-versus-host disease; inflammatory
CC   conditions e.g. ischaemia-reperfusion injury and arthritis;
CC   hyperproliferative disorders e.g. cancer and purpura; infectious disease
CC   e.g. viral infection and bacterial infection. The polynucleotide or
CC   protein can be used to regenerate damaged tissue e.g. ulcers and
CC   Alzheimer's disease. The present sequence represents the amino acid
CC   sequence of a novel human secreted protein. Note: The sequence data for
CC   this patent did not form part of the printed specification but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030065160
XX
SQ   Sequence 257 AA;

  Query Match             40.3%;  Score 1230;  DB 6;  Length 257;
  Best Local Similarity   93.1%;  Pred. No. 3.3e-110;
  Matches  241;  Conservative    3;  Mismatches    9;  Indels    6;  Gaps    1;

Qy        160 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 60

Qy        220 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVA 279
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVXSDSKRESVCNKADAESEPVA 120

Qy        280 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 180

Qy        340 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIF 399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT----- 235

Qy        400 QEDTIPRTQIERRKTSLYF 418
              ||||||    :::    |
Db        236 -EDTIPRNTDRKKENKPVF 253


RESULT 10
ABO26205
ID   ABO26205 standard; protein; 257 AA.
XX
AC   ABO26205;
XX
DT   10-SEP-2003  (first entry)
XX
DE   Human protein from novel secreted protein gene 168.
XX
KW   Human; secreted protein; precerebellin-like protein;
KW   neurodegenerative disorder; behavioural disorder; Alzheimer's disease;
KW   Parkinson's disease; Huntington's disease; schizophrenia; mania;
KW   dementia; paranoia; psychosis; autism; immune disorder; infection;
KW   inflammation; allergy; liver disorder; hepatoblastoma; jaundice;
KW   hepatitis; immunological disorder; AIDS; leukaemia; rheumatoid arthritis;
KW   sepsis; acne; psoriasis; cancer.
XX
OS   Homo sapiens.
XX
PN   US6525174-B1.
XX
PD   25-FEB-2003.
XX
PF   04-DEC-1998;   98US-00205258.
XX
PR   06-JUN-1997;   97US-0048875P.
PR   06-JUN-1997;   97US-0048876P.
PR   06-JUN-1997;   97US-0048877P.
PR   06-JUN-1997;   97US-0048878P.
PR   06-JUN-1997;   97US-0048880P.
PR   06-JUN-1997;   97US-0048881P.
PR   06-JUN-1997;   97US-0048882P.
PR   06-JUN-1997;   97US-0048883P.
PR   06-JUN-1997;   97US-0048884P.
PR   06-JUN-1997;   97US-0048885P.
PR   06-JUN-1997;   97US-0048892P.
PR   06-JUN-1997;   97US-0048893P.
PR   06-JUN-1997;   97US-0048894P.
PR   06-JUN-1997;   97US-0048895P.
PR   06-JUN-1997;   97US-0048896P.
PR   06-JUN-1997;   97US-0048897P.
PR   06-JUN-1997;   97US-0048898P.
PR   06-JUN-1997;   97US-0048899P.
PR   06-JUN-1997;   97US-0048900P.
PR   06-JUN-1997;   97US-0048901P.
PR   06-JUN-1997;   97US-0048915P.
PR   06-JUN-1997;   97US-0048916P.
PR   06-JUN-1997;   97US-0048917P.
PR   06-JUN-1997;   97US-0048949P.
PR   06-JUN-1997;   97US-0048962P.
PR   06-JUN-1997;   97US-0048963P.
PR   06-JUN-1997;   97US-0048964P.
PR   06-JUN-1997;   97US-0048970P.
PR   06-JUN-1997;   97US-0048971P.
```

PR 06-JUN-1997; 97US-0048972P.
PR 06-JUN-1997; 97US-0048974P.
PR 06-JUN-1997; 97US-0049019P.
PR 06-JUN-1997; 97US-0049020P.
PR 06-JUN-1997; 97US-0049373P.
PR 06-JUN-1997; 97US-0049374P.
PR 06-JUN-1997; 97US-0049375P.
PR 05-SEP-1997; 97US-0057584P.
PR 05-SEP-1997; 97US-0057627P.
PR 05-SEP-1997; 97US-0057628P.
PR 05-SEP-1997; 97US-0057629P.
PR 05-SEP-1997; 97US-0057634P.
PR 05-SEP-1997; 97US-0057635P.
PR 05-SEP-1997; 97US-0057642P.
PR 05-SEP-1997; 97US-0057643P.
PR 05-SEP-1997; 97US-0057644P.
PR 05-SEP-1997; 97US-0057645P.
PR 05-SEP-1997; 97US-0057646P.
PR 05-SEP-1997; 97US-0057647P.
PR 05-SEP-1997; 97US-0057648P.
PR 05-SEP-1997; 97US-0057649P.
PR 05-SEP-1997; 97US-0057650P.
PR 05-SEP-1997; 97US-0057651P.
PR 05-SEP-1997; 97US-0057654P.
PR 05-SEP-1997; 97US-0057661P.
PR 05-SEP-1997; 97US-0057662P.
PR 05-SEP-1997; 97US-0057666P.
PR 05-SEP-1997; 97US-0057667P.
PR 05-SEP-1997; 97US-0057668P.
PR 05-SEP-1997; 97US-0057760P.
PR 05-SEP-1997; 97US-0057761P.
PR 05-SEP-1997; 97US-0057762P.
PR 05-SEP-1997; 97US-0057763P.
PR 05-SEP-1997; 97US-0057764P.
PR 05-SEP-1997; 97US-0057765P.
PR 05-SEP-1997; 97US-0057769P.
PR 05-SEP-1997; 97US-0057770P.
PR 05-SEP-1997; 97US-0057771P.
PR 05-SEP-1997; 97US-0057774P.
PR 05-SEP-1997; 97US-0057775P.
PR 05-SEP-1997; 97US-0057776P.
PR 05-SEP-1997; 97US-0057777P.
PR 05-SEP-1997; 97US-0057778P.
PR 18-DEC-1997; 97US-0070923P.
PR 04-JUN-1998; 98WO-US011422.
PR 15-JUL-1998; 98US-0092921P.
PR 30-JUL-1998; 98US-0094657P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Young P, Greene JM, Ferrie AM, Ruben SM, Rosen CA, Hu J;
PI Olsen HS, Ebner R, Brewer LA, Moore PA, Shi Y, Florence C;
PI Florence K, Lafleur DW, Ni J, Fan P, Wei Y, Fischer CL, Soppet DR;
PI Li Y, Zeng Z, Kyaw H, Yu G, Feng P, Dillon PJ, Endress GA;
PI Carter KC;
XX
DR WPI; 2003-511926/48.
DR N-PSDB; ACD44672.
XX
PT New precerebellin-like protein, useful for diagnosing or treating
PT neurodegenerative and behavioral disorders, immune disorders, liver
PT disorders, and cancer.
XX
PS Disclosure; SEQ ID NO 416; 156pp; English.
XX
CC The invention relates to an isolated protein comprising amino acid
CC residues 33-205 or 1-205 of a a novel human secreted protein appearing as
CC ABO26252. The protein is encoded by one of 238 disclosed cDNA sequences
CC encoding 238 secreted proteins. ABO26252 is a precerebellin-like protein.
CC Also included are a composition comprising the protein and a carrier and
CC an isolated protein produced by expressing the protein cited above by a
CC cell, and recovering the protein. The proteins are useful for diagnosing
CC or treating neurodegenerative and behavioural disorders (e.g. Alzheimer's
CC disease, Parkinson's disease, Huntington's disease, schizophrenia, mania,
CC dementia, paranoia, psychoses or autism), immune disorders (e.g.
CC infection, inflammation, allergy), liver disorders (e.g. hepatoblastoma,
CC jaundice, hepatitis), immunological disorders (e.g. AIDS, leukaemia,
CC rheumatoid arthritis, sepsis, acne, psoriasis) and cancer. The present
CC sequence is one of the 238 disclosed novel secreted proteins. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at:- seqdata.uspto.gov/sequence.html?DocID=6525174B1
XX
SQ Sequence 257 AA;

Query Match 40.3%; Score 1230; DB 7; Length 257;
Best Local Similarity 93.1%; Pred. No. 3.3e-110;
Matches 241; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

```
Qy        160 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 60

Qy        220 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVA 279
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVXSDSKRESVCNKADAESEPVA 120

Qy        280 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 180

Qy        340 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIF 399
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT----- 235

Qy        400 QEDTIPRTQIERRKTSLYF 418
              ||||||    :::    |
Db        236 -EDTIPRNTDRKKENKPVF 253
```

RESULT 11
ADD89916
ID ADD89916 standard; protein; 202 AA.
XX
AC ADD89916;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human 5-methylcytosine DNA glycosylase N-terminal deletion mutant.
XX
KW Human; 5-methylcytosine DNA glycosylase; enzyme; CpG; mutant; mutein.
XX
OS Homo sapiens.
XX
PN WO2003078593-A2.
XX
PD 25-SEP-2003.
XX
PF 14-MAR-2003; 2003WO-US007933.
XX
PR 15-MAR-2002; 2002US-0364689P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Lofton-Day CE, Day JK;
XX
DR WPI; 2003-779127/73.
XX
PT Labeling methylated or methylatable CpG sequences, useful e.g. for
PT diagnostic detection of altered methylation, comprises replacing
PT methylated cytosine by labeled cytosine.
XX
PS Claim 11; Page 71-72; 73pp; English.
XX
CC The present sequence is the protein sequence of an N-terminal deletion
CC mutant of human 5-methylcytosine DNA glycosylase (5-MCDG), in which amino

```
CC   acid residue 1 corresponds to amino acid 379 of the full-length protein
CC   ADD89906. 5-MCDG acts by cleaving glycosylic bonds at methylated CpG
CC   sites of DNA, removing 5-methylcytosine from the DNA backbone as a free
CC   base. The N-terminal deletion mutant shows enhanced deglycosylase
CC   specificity towards CpG dinucleotide sequences. Human 5-MCDG can be used
CC   in a claimed method for labelling CpG sequences corresponding to
CC   methylated CpG sequences in an isolated DNA sample. The method comprises:
CC   digesting the genomic DNA with a restriction endonuclease to produce
CC   genomic DNA fragments; treating the genomic DNA fragments with 5-MCDG
CC   such that one or more 5-methylcytosine bases are removed to produce
CC   abasic genomic DNA fragments; and treating these abasic genomic DNA
CC   fragments with base excision repair enzymes in the presence of labelled
CC   dCTP such that 5-methylcytosine removed from the genomic DNA fragments by
CC   5-MCDG is replaced by labelled cytosine in the one or more corresponding
CC   positions of the abasic genomic DNA fragments to produce labelled genomic
CC   DNA fragments, so that specific labelling of CpG sequences corresponding
CC   to methylated CpG sequences is achieved. The 5-MCDG is also used in a
CC   claimed method for comparing CpG methylation status, extent or pattern
CC   between or among reference and test genomic DNA samples, and in a claimed
CC   method for labelling potentially-methylatable CpG sequences in CpG-
CC   containing genomic DNA fragments. The methods are used to identify
CC   methylated and/or potentially methylatable CpG dinucleotides in genomic
CC   DNA, including comparison of methylation pattern between healthy and
CC   diseased samples, for diagnosis.
XX
SQ   Sequence 202 AA;

  Query Match             36.2%;  Score 1106;  DB 7;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-98;
  Matches  202;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        379 GSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKW 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKW 60

Qy        439 TPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRD 498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRD 120

Qy        499 VSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHP 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHP 180

Qy        559 EDHKLNKYHDWLWENHEKLSLS 580
              ||||||||||||||||||||||
Db        181 EDHKLNKYHDWLWENHEKLSLS 202


RESULT 12
ADD89908
ID   ADD89908 standard; protein; 416 AA.
XX
AC   ADD89908;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Chicken 5-methylcytosine DNA glycosylase.
XX
KW   Chicken; 5-methylcytosine DNA glycosylase; enzyme; CpG.
XX
OS   Gallus gallus.
XX
PN   WO2003078593-A2.
XX
PD   25-SEP-2003.
XX
PF   14-MAR-2003; 2003WO-US007933.
XX
PR   15-MAR-2002; 2002US-0364689P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Lofton-Day CE,  Day JK;
XX
DR   WPI; 2003-779127/73.
DR   N-PSDB; ADD89907.
XX
PT   Labeling methylated or methylatable CpG sequences, useful e.g. for
PT   diagnostic detection of altered methylation, comprises replacing
PT   methylated cytosine by labeled cytosine.
XX
PS   Claim 11; Page 56-58; 73pp; English.
XX
CC   The present sequence is the protein sequence of chicken 5-methylcytosine
CC   DNA glycosylase (5-MCDG). The enzyme acts by cleaving glycosylic bonds at
CC   methylated CpG sites of DNA, removing 5-methylcytosine from the DNA
CC   backbone as a free base. Chicken 5-MCDG can be used in a claimed method
CC   for labelling CpG sequences corresponding to methylated CpG sequences in
CC   an isolated DNA sample. The method comprises: digesting the genomic DNA
CC   with a restriction endonuclease to produce genomic DNA fragments;
CC   treating the genomic DNA fragments with 5-MCDG such that one or more 5-
CC   methylcytosine bases are removed to produce abasic genomic DNA fragments;
CC   and treating these abasic genomic DNA fragments with base excision repair
CC   enzymes in the presence of labelled dCTP such that 5-methylcytosine
CC   removed from the genomic DNA fragments by 5-MCDG is replaced by labelled
CC   cytosine in the one or more corresponding positions of the abasic genomic
CC   DNA fragments to produce labelled genomic DNA fragments, so that specific
CC   labelling of CpG sequences corresponding to methylated CpG sequences is
CC   achieved. The 5-MCDG is also used in a claimed method for comparing CpG
CC   methylation status, extent or pattern between or among reference and test
CC   genomic DNA samples, and in a claimed method for labelling potentially-
CC   methylatable CpG sequences in CpG-containing genomic DNA fragments. The
CC   methods are used to identify methylated and/or potentially methylatable
CC   CpG dinucleotides in genomic DNA, including comparison of methylation
CC   pattern between healthy and diseased samples, for diagnosis.
XX
SQ   Sequence 416 AA;

  Query Match             28.8%;  Score 880.5;  DB 7;  Length 416;
  Best Local Similarity   78.2%;  Pred. No. 6.3e-76;
  Matches  161;  Conservative   20;  Mismatches   21;  Indels    4;  Gaps    2;

Qy        378 RGSEMDNNCS--PTRKDFTGEKIFQ--EDTIPRTQIERRKTSLYFSSKYNKEALSPPRRK 433
              | |  | : |    :| ||  :  :  |:: ||||::||||| |||||:||||||||||
Db        210 RDSAADGDVSWPSDKKSFTAVQAPRGTEESAPRTQVDRRKTSPYFSSKYSKEALSPPRRK 269

Qy        434 AFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVART 493
              ||:||||||||||||||||||||||||||||:||||||||||:|| |||| |||||
Db        270 AFRKWTPPRSPFNLVQETLFHDPWKLLIATIFLNKTSGKMAIPVLWEFLRKYPSPEVART 329

Qy        494 ADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEW 553
              |||:::||||:||||| ||||||:||||||| ||||||||||||||||||||||||||||
Db        330 ADWKEMSELLRPLGLYALRAKTIIKFSDEYLNKQWKYPIELHGIGKYGNDSYRIFCVNEW 389

Qy        554 KQVHPEDHKLNKYHDWLWENHEKLSL 579
              |:| |:||||| || ||||||||||:
Db        390 KEVQPQDHKLNIYHAWLWENHEKLSV 415


RESULT 13
ADD89917
ID   ADD89917 standard; protein; 147 AA.
XX
AC   ADD89917;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human 5-methylcytosine DNA glycosylase N-terminal deletion mutant.
XX
KW   Human; 5-methylcytosine DNA glycosylase; enzyme; CpG; mutant; mutein.
XX
OS   Homo sapiens.
XX
PN   WO2003078593-A2.
XX
PD   25-SEP-2003.
```

XX
PF 14-MAR-2003; 2003WO-US007933.
XX
PR 15-MAR-2002; 2002US-0364689P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Lofton-Day CE, Day JK;
XX
DR WPI; 2003-779127/73.
XX
PT Labeling methylated or methylatable CpG sequences, useful e.g. for
PT diagnostic detection of altered methylation, comprises replacing
PT methylated cytosine by labeled cytosine.
XX
PS Claim 11; Page 73; 73pp; English.
XX
CC The present sequence is the protein sequence of an N-terminal deletion
CC mutant of human 5-methylcytosine DNA glycosylase (5-MCDG), in which amino
CC acid residue 1 corresponds to amino acid 434 of the full-length protein
CC ADD89906. 5-MCDG acts by cleaving glycosylic bonds at methylated CpG
CC sites of DNA, removing 5-methylcytosine from the DNA backbone as a free
CC base. The N-terminal deletion mutant shows enhanced deglycosylase
CC specificity towards CpG dinucleotide sequences. Human 5-MCDG can be used
CC in a claimed method for labelling CpG sequences corresponding to
CC methylated CpG sequences in an isolated DNA sample. The method comprises:
CC digesting the genomic DNA with a restriction endonuclease to produce
CC genomic DNA fragments; treating the genomic DNA fragments with 5-MCDG
CC such that one or more 5-methylcytosine bases are removed to produce
CC abasic genomic DNA fragments; and treating these abasic genomic DNA
CC fragments with base excision repair enzymes in the presence of labelled
CC dCTP such that 5-methylcytosine removed from the genomic DNA fragments by
CC 5-MCDG is replaced by labelled cytosine in the one or more corresponding
CC positions of the abasic genomic DNA fragments to produce labelled genomic
CC DNA fragments, so that specific labelling of CpG sequences corresponding
CC to methylated CpG sequences is achieved. The 5-MCDG is also used in a
CC claimed method for comparing CpG methylation status, extent or pattern
CC between or among reference and test genomic DNA samples, and in a claimed
CC method for labelling potentially-methylatable CpG sequences in CpG-
CC containing genomic DNA fragments. The methods are used to identify
CC methylated and/or potentially methylatable CpG dinucleotides in genomic
CC DNA, including comparison of methylation pattern between healthy and
CC diseased samples, for diagnosis.
XX
SQ Sequence 147 AA;

Query Match 26.7%; Score 816; DB 7; Length 147;
Best Local Similarity 100.0%; Pred. No. 2e-70;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        434 AFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVART 493
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVART 60

Qy        494 ADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEW 553
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEW 120

Qy        554 KQVHPEDHKLNKYHDWLWENHEKLSLS 580
              ||||||||||||||||||||||||||
Db        121 KQVHPEDHKLNKYHDWLWENHEKLSLS 147
```

RESULT 14
AAE22568
ID AAE22568 standard; peptide; 68 AA.
XX
AC AAE22568;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human MBD4 methyl CpG binding motif.
XX
KW Gene expression; cellular chromatin; methyl CpG binding domain; cancer;
KW localisation domain; diabetic retinopathy; ischaemia; HIV infection;
KW human immuno deficiency virus; macular degeneration; vascular disease;
KW rheumatoid arthritis; psoriasis; Alzheimer's disease; muscular dystrophy;
KW sickle cell anaemia; stroke; neurodegenerative disease; cystic fibrosis;
KW gene therapy; cytostatic; antidiabetic; ophthalmological; vasotropic;
KW neuroprotective; nootropic; cerebroprotective; antibacterial; antifungal;
KW antiviral; human; MBD4 methyl CpG binding motif.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1. .3
FT /label= Beta1_helix
FT Region 10. .15
FT /label= Beta2_helix
FT Region 16. .26
FT /label= L1_loop
FT Region 27. .32
FT /label= Beta3_helix
FT Region 36. .38
FT /label= Beta4_helix
FT Region 42. .48
FT /label= Alpha1_helix
FT Region 49. .60
FT /label= L2_loop
FT Region 61. .68
FT /label= Hairpin_loop
XX
PN WO200226960-A2.
XX
PD 04-APR-2002.
XX
PF 28-SEP-2001; 2001WO-US042377.
XX
PR 29-SEP-2000; 2000US-0236884P.
XX
PA (SANG-) SANGAMO BIOSCIENCES INC.
XX
PI Wolffe AP, Urnov F, Lai A, Raschke E;
XX
DR WPI; 2002-372124/40.
XX
PT Compartmentalizing a region of interest in cellular chromatin which
PT facilitates the modulation of the expression of a gene comprises
PT contacting the gene with a composition comprising a localization domain
PT and a DNA binding domain.
XX
PS Example 2; Fig 1A; 85pp; English.
XX
CC The present invention relates to methods and compositions for regulating
CC gene expression. In particular the method of compartmentalising a region
CC of interest in cellular chromatin comprises contacting the region of
CC interest with a composition that binds to a binding site in cellular
CC chromatin, where the binding site is in a gene of interest and the
CC composition comprises a localisation domain (e.g., methyl CpG binding
CC domain obtained from MECP2, MBD1, MBD2, MBD3, dMBD-like and dMBD-like
CC delta) and a DNA binding domain (or functional fragment). The method is
CC useful for compartmentalising a region of interest in cellular chromatin
CC which facilitates the modulation of the expression of a gene using a
CC fusion molecule comprising a DNA binding domain and a localisation domain
CC that binds to the chromatin. The fusion molecules or polypeptides can be
CC used to prepare pharmaceutical compositions to prevent or treat cancer,
CC ischaemia, diabetic retinopathy, macular degeneration, HIV infection,
CC rheumatoid arthritis, psoriasis, sickle cell anaemia, vascular disease,
CC Alzheimer's disease, muscular dystrophy, neurodegenerative diseases,
CC cystic fibrosis, stroke, bacterial, viral or fungal infections. Sequences
CC of the invention are also used in gene therapy. The present sequence is
CC human MBD4 methyl CpG binding motif. This sequence is used in the
CC exemplification of the invention
XX
SQ Sequence 68 AA;

```
  Query Match             11.7%;  Score 357;  DB 5;  Length 68;
  Best Local Similarity   98.5%;  Pred. No. 2.1e-26;
  Matches   67;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         81 ECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPED 140
               ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db          1 ECRKSVPCGWERVVKQRLFGKTAGRFDVYFYSPQGLKFRSKSSLANYLHKNGETSLKPED 60

Qy        141 FDFTVLSK 148
               ||||||||
Db         61 FDFTVLSK 68


RESULT 15
ABB51147
ID   ABB51147 standard; protein; 50 AA.
XX
AC   ABB51147;
XX
DT   07-FEB-2002  (first entry)
XX
DE   Human secreted protein encoded by gene 168 SEQ ID NO:1100.
XX
KW   Human; secreted protein; immunomodulatory; antisclerotic; anti-HIV;
KW   dermatological; immunosuppressive; antiinflammatory; immunostimulant;
KW   cytostatic; cardiant; vascular; anti-angiogenic; ophthalmological;
KW   neuroprotective; nootropic; anticonvulsant; antialzheimers; vulnerary;
KW   antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;
KW   multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;
KW   human immunodeficiency virus; hyperproliferative disorder; wound healing;
KW   Gaucher's disease; cardiovascular disease; Scimitar syndrome; chemotaxis;
KW   Chaga's cardiomyopathy; coronary arteriosclerosis; angiogenic disorder;
KW   corneal graft neovascularisation; diabetic retinopathy; regeneration;
KW   neurological disorder; Huntington's chorea; Alzheimer's disease;
KW   Parkinson's disease; infectious disease; chromosome 3.
XX
OS   Homo sapiens.
XX
PN   WO200162891-A2.
XX
PD   30-AUG-2001.
XX
PF   21-FEB-2001; 2001WO-US005614.
XX
PR   24-FEB-2000; 2000US-0184836P.
PR   29-MAR-2000; 2000US-0193170P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ni J,  Ebner R,  Lafleur DW,  Moore PA,  Olsen HS,  Rosen CA;
PI   Ruben SM,  Soppet DR,  Young PE,  Shi Y,  Florence KA,  Wei Y;
PI   Florence C,  Hu J,  Li Y,  Kyaw H,  Fischer CL,  Ferrie AM,  Fan P;
PI   Feng P,  Endress GA,  Dillon PJ,  Carter KC,  Brewer LA,  Yu G,  Zeng Z;
PI   Greene JM;
XX
DR   WPI; 2001-625724/72.
XX
PT   Nucleic acids encoding 207 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating, e.g. cancers, Parkinson's disease
PT   and diabetic retinopathy.
XX
PS   Disclosure; Page 366; 1533pp; English.
XX
CC   ABB50301 to ABB51287 and ABA83194 to ABA83441 represent human secreted
CC   proteins (I) and polynucleotide (II) sequences. (I) and (II) have various
CC   activities based on the tissues and cells the genes are expressed in.
CC   Example of these activities include: immunomodulatory; antisclerotic;
CC   dermatological; immunosuppressive; antiinflammatory; immunostimulant;
CC   anti-HIV; cytostatic; cardiant; anti-angiogenic; ophthalmological;
CC   neuroprotective; nootropic; anticonvulsant; antialzheimers; vascular;
CC   antiparkinsonian; antimicrobial; and vulnerary. (I) and (II) can be used
CC   in gene therapy and vaccine production. (I) and (II) can be used in the
CC   prevention, diagnosis and treatment of immune disorders (e.g. multiple
CC   sclerosis, systemic lupus erythematosus and human immunodeficiency virus
CC   (HIV) infections), hyperproliferative disorders (e.g. cancers and
CC   Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,
CC   Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic
CC   disorders (e.g. corneal graft neovascularisation and diabetic
CC   retinopathy), neurological disorders (e.g. Huntington's chorea,
CC   Alzheimer's disease and Parkinson's disease), infectious diseases and/or
CC   for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to
CC   ABA83193 and ABB50300 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 50 AA;

  Query Match             8.3%;  Score 254;  DB 4;  Length 50;
  Best Local Similarity   98.0%;  Pred. No. 1.3e-16;
  Matches   49;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        110 FISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCS 159
              | ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FSSPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCS 50


Search completed: August 22, 2005, 10:05:09
Job time : 172 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 22, 2005, 10:02:09 ; Search time 163 Seconds
(without alignments)
1393.374 Million cell updates/sec

Title: US-10-629-951-2
Perfect score: 3055
Sequence: 1 MGTTGLESLSLGDRGAAPTV.....HKLNKYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1759131 seqs, 391586102 residues

Total number of hits satisfying chosen parameters: 1759131

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3055	100.0	580	14	US-10-389-853-2	Sequence 2, Appli
2	3055	100.0	580	15	US-10-629-951-2	Sequence 2, Appli
3	3011	98.6	574	15	US-10-629-951-24	Sequence 24, Appl
4	1639	53.6	307	14	US-10-389-853-11	Sequence 11, Appl
5	1326	43.4	384	15	US-10-629-951-29	Sequence 29, Appl
6	1230	40.3	257	10	US-09-933-767-416	Sequence 416, App
7	1230	40.3	257	14	US-10-004-860-416	Sequence 416, App
8	1230	40.3	257	14	US-10-023-282-416	Sequence 416, App
9	1106	36.2	202	14	US-10-389-853-12	Sequence 12, Appl
10	880.5	28.8	416	14	US-10-389-853-4	Sequence 4, Appli
11	816	26.7	147	14	US-10-389-853-13	Sequence 13, Appl
12	697	22.8	126	15	US-10-629-951-32	Sequence 32, Appl
13	616	20.2	119	15	US-10-629-951-30	Sequence 30, Appl
14	443	14.5	85	15	US-10-629-951-37	Sequence 37, Appl
15	357	11.7	68	10	US-09-967-869A-7	Sequence 7, Appli
16	357	11.7	68	20	US-11-045-828-7	Sequence 7, Appli
17	330	10.8	443	16	US-10-739-930-5943	Sequence 5943, Ap
18	300.5	9.8	382	16	US-10-425-115-299452	Sequence 299452,
19	281	9.2	185	16	US-10-767-701-33975	Sequence 33975, A
20	254	8.3	50	10	US-09-933-767-1100	Sequence 1100, Ap
21	254	8.3	50	14	US-10-004-860-1100	Sequence 1100, Ap
22	254	8.3	50	14	US-10-023-282-1100	Sequence 1100, Ap
23	209	6.8	476	16	US-10-408-765A-266	Sequence 266, App
24	208.5	6.8	560	16	US-10-475-681-7	Sequence 7, Appli
25	208.5	6.8	561	16	US-10-475-681-3	Sequence 3, Appli
26	192	6.3	132	15	US-10-629-951-31	Sequence 31, Appl
27	186	6.1	68	10	US-09-967-869A-5	Sequence 5, Appli
28	186	6.1	68	20	US-11-045-828-5	Sequence 5, Appli
29	152	5.0	105	16	US-10-437-963-148365	Sequence 148365,
30	145	4.7	1301	15	US-10-369-493-1644	Sequence 1644, Ap
31	145	4.7	3418	15	US-10-392-113-42	Sequence 42, Appl
32	145	4.7	3418	15	US-10-634-574-1	Sequence 1, Appli
33	145	4.7	3418	16	US-10-408-765A-178	Sequence 178, App
34	137	4.5	560	14	US-10-205-841-12	Sequence 12, Appl
35	137	4.5	1713	18	US-10-840-512-116	Sequence 116, App
36	136.5	4.5	409	16	US-10-425-115-250086	Sequence 250086,
37	136.5	4.5	1359	17	US-10-732-923-8708	Sequence 8708, Ap
38	135.5	4.4	282	10	US-09-967-869A-17	Sequence 17, Appl
39	135.5	4.4	282	20	US-11-045-828-17	Sequence 17, Appl
40	135.5	4.4	755	16	US-10-473-127-548	Sequence 548, App
41	135.5	4.4	755	16	US-10-473-127-550	Sequence 550, App
42	135.5	4.4	755	16	US-10-473-127-553	Sequence 553, App
43	133.5	4.4	1359	17	US-10-732-923-8707	Sequence 8707, Ap
44	132	4.3	578	9	US-09-821-839-2	Sequence 2, Appli
45	132	4.3	1819	15	US-10-335-977-7981	Sequence 7981, Ap

ALIGNMENTS

RESULT 1
US-10-389-853-2
; Sequence 2, Application US/10389853
; Publication No. US20030180779A1
; GENERAL INFORMATION:
; APPLICANT: Lofton-Day, Cathy E.
; APPLICANT: Day, John K.
; TITLE OF INVENTION: Discovery and Diagnostic Methods Using 5-Methylcytosine DNA Glyco
; FILE REFERENCE: 47675-36
; CURRENT APPLICATION NUMBER: US/10/389,853
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 60/364,689
; PRIOR FILING DATE: 2002-03-15
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-389-853-2

Query Match 100.0%; Score 3055; DB 14; Length 580;
Best Local Similarity 100.0%; Pred. No. 8.2e-236;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
```

```
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580

Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
```

RESULT 2
US-10-629-951-2
; Sequence 2, Application US/10629951
; Publication No. US20040018550A1
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; TITLE OF INVENTION: Single-Nucleotide Polymorphisns
; FILE REFERENCE: FCCC 96-21
; CURRENT APPLICATION NUMBER: US/10/629,951
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: US/09/629,222A
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-629-951-2

Query Match 100.0%; Score 3055; DB 15; Length 580;
Best Local Similarity 100.0%; Pred. No. 8.2e-236;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580

Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
```

RESULT 3
US-10-629-951-24
; Sequence 24, Application US/10629951
; Publication No. US20040018550A1
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; TITLE OF INVENTION: Single-Nucleotide Polymorphisns
; FILE REFERENCE: FCCC 96-21
; CURRENT APPLICATION NUMBER: US/10/629,951
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: US/09/629,222A
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-629-951-24

Query Match 98.6%; Score 3011; DB 15; Length 574;
Best Local Similarity 99.0%; Pred. No. 2.7e-232;
Matches 574; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

```
Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
```

```
Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||      ||||||||||||||||||||
Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT------EDTIPRTQIERRKTSLYFSS 414

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        415 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 474

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        475 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 534

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              |||||||||||||||||||||||||||||||||||||||
Db        535 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 574
```

RESULT 4
US-10-389-853-11
; Sequence 11, Application US/10389853
; Publication No. US20030180779A1
; GENERAL INFORMATION:
; APPLICANT: Lofton-Day, Cathy E.
; APPLICANT: Day, John K.
; TITLE OF INVENTION: Discovery and Diagnostic Methods Using 5-Methylcytosine DNA Glycc
; FILE REFERENCE: 47675-36
; CURRENT APPLICATION NUMBER: US/10/389,853
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 60/364,689
; PRIOR FILING DATE: 2002-03-15
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: NON_TER
; LOCATION: (1)..(1)
; OTHER INFORMATION: functional N-terminal deletion mutant of hMBD4 (SEQ ID NO:2); aa
; OTHER INFORMATION: no. 1 corresponds to aa no. 274 of hMBD4; mutant shows enhanced d
; OTHER INFORMATION: eglycosylase specificity towards CpG dinucleotide sequences; see
; OTHER INFORMATION: Zhu et al. Nuc. Acid Res. 28:4157-4165, 2000.
US-10-389-853-11

Query Match 53.6%; Score 1639; DB 14; Length 307;
Best Local Similarity 100.0%; Pred. No. 1e-122;
Matches 307; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        274 ESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGII 333
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ESEPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGII 60

Qy        334 NKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDF 393
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NKFCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDF 120

Qy        394 TGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLF 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLF 180

Qy        454 HDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRA 513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRA 240

Qy        514 KTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWEN 573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWEN 300

Qy        574 HEKLSLS 580
              |||||||
Db        301 HEKLSLS 307
```

RESULT 5
US-10-629-951-29
; Sequence 29, Application US/10629951
; Publication No. US20040018550A1
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; TITLE OF INVENTION: Single-Nucleotide Polymorphisns
; FILE REFERENCE: FCCC 96-21
; CURRENT APPLICATION NUMBER: US/10/629,951
; CURRENT FILING DATE: 2003-07-29
; PRIOR APPLICATION NUMBER: US/09/629,222A
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-629-951-29

Query Match 43.4%; Score 1326; DB 15; Length 384;
Best Local Similarity 54.3%; Pred. No. 1.7e-97;
Matches 283; Conservative 29; Mismatches 71; Indels 138; Gaps 4;

```
Qy         36 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVK 95
              |||:|: |  |||| : ::|  ||| : ||||| ||    :|  ||  | ||||||||||
Db          1 KEDIAVGLGGVGEDGKDLVI--SSERSSLLQEPTAST-LSSTTATEGHKPVPCGWERVVK 57

Qy         96 QRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRY 155
              ||| |||||:|||||||||||||||| |||||| ||||| ||||||:|||| |  |    |
Db         58 QRLSGKTAGKFDVYFISPQGLKFRSKRSLANYLLKNGETFLKPEDFNFTVLPKGSINPGY 117

Qy        156 KDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLK 215
              |  |:|||||    |::: |  ||:|||| | ||   || :||  || ||||  |  |||:
Db        118 KHQSLAALTSLQPNETDVSKQNLKTRSKWKTDVLPLPSGTSESPESSGLSNSNSACLLLR 177

Qy        216 EDEGVDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAES 275
              |   : ||:  | || | |||:|||   :|||: ||||      | : || |
Db        178 EHRDIQDVDSEKRRKSKRKVTVLKGTASQKTKQKCRKSLLESTQRNRKRAS--------- 228

Qy        276 EPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINK 335

Db        229 ------------------------------------------------------------ 228

Qy        336 FCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTG 395

Db        229 ------------------------------------------------------------ 228

Qy        396 EKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHD 455
              ||:|||||:|:|||||||||||||||||||:|||||||||||||||| ||||
Db        229 -----EDSIPRTQVEKRKTSLYFSSKYNKEALSPPRRKSFKKWTPPRSPFNLVQEILFHD 283
```

```
Qy        456 PWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKT 515
              ||||||||||||||||||||||||:||||||||||| ||||||||||||||||||||
Db        284 PWKLLIATIFLNRTSGKMAIPVLWEFLEKYPSAEVARAADWRDVSELLKPLGLYDLRAKT 343

Qy        516 IVKFSDEYLTKQWKYPIELHGIG-KYGNDSYRIFCVNEWKQ 555
              |:||||||||||:||||||||  ||||||||||||||||
Db        344 IIKFSDEYLTKQWRYPIELHGIWLKYGNDSYRIFCVNEWKQ 384


RESULT 6
US-09-933-767-416
; Sequence 416, Application US/09933767
; Publication No. US20030181692A1
; GENERAL INFORMATION:
;  APPLICANT: Ni et al.
;  TITLE OF INVENTION: 207 Human Secreted Proteins
;  FILE REFERENCE: PZ007P2
;  CURRENT APPLICATION NUMBER: US/09/933,767
;  CURRENT FILING DATE:  2001-08-22
;  PRIOR APPLICATION NUMBER: PCT/US01/05614
;  PRIOR FILING DATE: 2001-02-21
;  PRIOR APPLICATION NUMBER: 60/184,836
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/193,170
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: 09/205,258
;  PRIOR FILING DATE: 1998-12-04
;  PRIOR APPLICATION NUMBER: PCT/US98/11422
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/048,885
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/049,375
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,881
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,880
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,896
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/049,020
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,876
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,895
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,884
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,894
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,971
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,964
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,882
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,899
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,893
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,900
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,901
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,892
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,915
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/049,019
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,970
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,972
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,916
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/049,373
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,875
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/049,374
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,917
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,949
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,974
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,883
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,897
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,898
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,962
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,963
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,877
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/048,878
;  PRIOR FILING DATE: 1997-06-06
;  PRIOR APPLICATION NUMBER: 60/068,054
;  PRIOR FILING DATE: 1997-12-18
;  PRIOR APPLICATION NUMBER: 60/068,064
;  PRIOR FILING DATE: 1997-12-18
;  PRIOR APPLICATION NUMBER: 60/068,053
;  PRIOR FILING DATE: 1997-12-18
;  PRIOR APPLICATION NUMBER: 60/070,923
;  PRIOR FILING DATE: 1997-12-18
;  PRIOR APPLICATION NUMBER: 60/073,160
;  PRIOR FILING DATE: 1998-01-30
;  PRIOR APPLICATION NUMBER: 60/073,159
;  PRIOR FILING DATE: 1998-01-30
;  PRIOR APPLICATION NUMBER: 60/073,165
;  PRIOR FILING DATE: 1998-01-30
;  PRIOR APPLICATION NUMBER: 60/073,164
;  PRIOR FILING DATE: 1998-01-30
;  PRIOR APPLICATION NUMBER: 60/085,925
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,921
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,923
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/085,922
;  PRIOR FILING DATE: 1998-05-18
;  PRIOR APPLICATION NUMBER: 60/092,921
;  PRIOR FILING DATE: 1998-07-15
;  PRIOR APPLICATION NUMBER: 60/094,657
;  PRIOR FILING DATE: 1998-07-30
;  NUMBER OF SEQ ID NOS: 1245
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 416
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (100)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-933-767-416

  Query Match             40.3%;  Score 1230;  DB 10;  Length 257;
  Best Local Similarity   93.1%;  Pred. No. 4.6e-90;
  Matches  241;  Conservative    3;  Mismatches    9;  Indels    6;  Gaps    1;

Qy        160 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 219
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 60

Qy        220 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVA 279
                 |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVXSDSKRESVCNKADAESEPVA 120

Qy        280 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 339
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 180

Qy        340 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIF 399
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT----- 235

Qy        400 QEDTIPRTQIERRKTSLYF 418
                  ||||||    :::    |
Db        236 -EDTIPRNTDRKKENKPVF 253


RESULT 7
US-10-004-860-416
; Sequence 416, Application US/10004860
; Publication No. US20030065160A1
; GENERAL INFORMATION:
;  APPLICANT: Young et al.
;  TITLE OF INVENTION: 207 Human Secreted Proteins
;  FILE REFERENCE: PZ007P1
;  CURRENT APPLICATION NUMBER: US/10/004,860
;  CURRENT FILING DATE:  2001-12-07
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 1227
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 416
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (100)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-004-860-416

  Query Match            40.3%;  Score 1230;  DB 14;  Length 257;
  Best Local Similarity   93.1%;  Pred. No. 4.6e-90;
  Matches  241;  Conservative    3;  Mismatches    9;  Indels    6;  Gaps    1;

Qy        160 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 219
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 60

Qy        220 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVA 279
                 |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVXSDSKRESVCNKADAESEPVA 120

Qy        280 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 339
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 180

Qy        340 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIF 399
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT----- 235

Qy        400 QEDTIPRTQIERRKTSLYF 418
                  ||||||    :::    |
Db        236 -EDTIPRNTDRKKENKPVF 253


RESULT 8
US-10-023-282-416
; Sequence 416, Application US/10023282
; Publication No. US20030092893A1
; GENERAL INFORMATION:
;  APPLICANT: Young et al.
;  TITLE OF INVENTION: 207 Human Secreted Proteins
;  FILE REFERENCE: PZ007P1
;  CURRENT APPLICATION NUMBER: US/10/023,282
;  CURRENT FILING DATE: 2001-12-20
;  EARLIER APPLICATION NUMBER: 09/205,258
;  EARLIER FILING DATE: 1998-12-04
;  EARLIER APPLICATION NUMBER: PCT/US98/11422
;  EARLIER FILING DATE: 1998-06-04
;  EARLIER APPLICATION NUMBER: 60/048,885
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,375
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,881
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,880
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,896
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,020
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,876
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,895
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,884
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,894
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,971
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,882
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,899
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,893
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,900
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,901
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,892
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,915
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,019
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,970
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,972
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,916
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,373
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,875
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,374
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,917
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,949
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,974
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,883
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,897
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,898
;  EARLIER FILING DATE: 1997-06-06
```

```
;  EARLIER APPLICATION NUMBER: 60/048,962
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,963
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,877
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,878
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/070,923
;  EARLIER FILING DATE: 1997-12-18
;  EARLIER APPLICATION NUMBER: 60/092,921
;  EARLIER FILING DATE: 1998-07-15
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  NUMBER OF SEQ ID NOS: 1227
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 416
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (100)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-023-282-416

  Query Match             40.3%;  Score 1230;  DB 14;  Length 257;
  Best Local Similarity   93.1%;  Pred. No. 4.6e-90;
  Matches  241;  Conservative    3;  Mismatches    9;  Indels    6;  Gaps    1;

Qy        160 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 60

Qy        220 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVA 279
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVXSDSKRESVCNKADAESEPVA 120

Qy        280 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 180

Qy        340 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIF 399
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT----- 235

Qy        400 QEDTIPRTQIERRKTSLYF 418
              ||||||    :::    |
Db        236 -EDTIPRNTDRKKENKPVF 253


RESULT 9
US-10-389-853-12
; Sequence 12, Application US/10389853
; Publication No. US20030180779A1
; GENERAL INFORMATION:
;  APPLICANT: Lofton-Day, Cathy E.
;  APPLICANT:  Day, John K.
;  TITLE OF INVENTION: Discovery and Diagnostic Methods Using 5-Methylcytosine DNA Glycc
;  FILE REFERENCE: 47675-36
;  CURRENT APPLICATION NUMBER: US/10/389,853
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: 60/364,689
;  PRIOR FILING DATE: 2002-03-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 202
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: NON_TER
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: functional N-terminal deletion mutant of hMBD4 (SEQ ID NO:2); aa
;   OTHER INFORMATION: no. 1 corresponds to aa no. 379 of hMBD4 protein; mutant shows en
;   OTHER INFORMATION: hanced deglycosylase specificity towards CpG dinucleotide sequenc
;   OTHER INFORMATION: es; see Zhu et al. Nuc. Acid Res. 28:4157-4165, 2000.
US-10-389-853-12

  Query Match             36.2%;  Score 1106;  DB 14;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-80;
  Matches  202;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        379 GSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKW 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKW 60

Qy        439 TPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRD 498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRD 120

Qy        499 VSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHP 558
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHP 180

Qy        559 EDHKLNKYHDWLWENHEKLSLS 580
              |||||||||||||||||||||||
Db        181 EDHKLNKYHDWLWENHEKLSLS 202


RESULT 10
US-10-389-853-4
; Sequence 4, Application US/10389853
; Publication No. US20030180779A1
; GENERAL INFORMATION:
;  APPLICANT: Lofton-Day, Cathy E.
;  APPLICANT:  Day, John K.
;  TITLE OF INVENTION: Discovery and Diagnostic Methods Using 5-Methylcytosine DNA Glycos
;  FILE REFERENCE: 47675-36
;  CURRENT APPLICATION NUMBER: US/10/389,853
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: 60/364,689
;  PRIOR FILING DATE: 2002-03-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Gallus gallus
US-10-389-853-4

  Query Match             28.8%;  Score 880.5;  DB 14;  Length 416;
  Best Local Similarity   78.2%;  Pred. No. 9.6e-62;
  Matches  161;  Conservative   20;  Mismatches   21;  Indels    4;  Gaps    2;

Qy        378 RGSEMDNNCS--PTRKDFTGEKIFQ--EDTIPRTQIERRKTSLYFSSKYNKEALSPPRRK 433
              | |  | : |    :| ||  :  :  |:: ||||::||||| |||||||:||||||||||
Db        210 RDSAADGDVSWPSDKKSFTAVQAPRGTEESAPRTQVDRRKTSPYFSSKYSKEALSPPRRK 269

Qy        434 AFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVART 493
              ||:||||||||||||||||||||||||||||:||||||||||:|| |||| |||||
Db        270 AFRKWTPPRSPFNLVQETLFHDPWKLLIATIFLNKTSGKMAIPVLWEFLRKYPSPEVART 329

Qy        494 ADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEW 553
              |||:::||||:||||| |||||:||||||| ||||||||||||||||||||||||||||||
Db        330 ADWKEMSELLRPLGLYALRAKTIIKFSDEYLNKQWKYPIELHGIGKYGNDSYRIFCVNEW 389

Qy        554 KQVHPEDHKLNKYHDWLWENHEKLSL 579
              |:| |:||||| || ||||||||||:
Db        390 KEVQPQDHKLNIYHAWLWENHEKLSV 415


RESULT 11
US-10-389-853-13
; Sequence 13, Application US/10389853
; Publication No. US20030180779A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Lofton-Day, Cathy E.
;  APPLICANT:  Day, John K.
;  TITLE OF INVENTION: Discovery and Diagnostic Methods Using 5-Methylcytosine DNA Glycc
;  FILE REFERENCE: 47675-36
;  CURRENT APPLICATION NUMBER: US/10/389,853
;  CURRENT FILING DATE:  2003-03-14
;  PRIOR APPLICATION NUMBER: 60/364,689
;  PRIOR FILING DATE: 2002-03-15
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
;   LENGTH: 147
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: NON_TER
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: functional N-terminal deletion mutant of hMBD4 (SEQ ID NO:2); aa
;   OTHER INFORMATION: no. 1 corresponds to aa no. 434 of hMBD4 protein; mutant shows en
;   OTHER INFORMATION: hanced deglycosylase specificity towards CpG dinucleotide sequenc
;   OTHER INFORMATION: es; see Zhu et al. Nuc. Acid Res. 28:4157-4165, 2000.
US-10-389-853-13

  Query Match             26.7%;  Score 816;  DB 14;  Length 147;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-57;
  Matches  147;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        434 AFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVART 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVART 60

Qy        494 ADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEW 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEW 120

Qy        554 KQVHPEDHKLNKYHDWLWENHEKLSLS 580
              ||||||||||||||||||||||||||||
Db        121 KQVHPEDHKLNKYHDWLWENHEKLSLS 147


RESULT 12
US-10-629-951-32
; Sequence 32, Application US/10629951
; Publication No. US20040018550A1
; GENERAL INFORMATION:
;  APPLICANT: Bellacosa, Alfonso
;  TITLE OF INVENTION: Methods for Detection of Transition
;  TITLE OF INVENTION:  Single-Nucleotide Polymorphisns
;  FILE REFERENCE: FCCC 96-21
;  CURRENT APPLICATION NUMBER: US/10/629,951
;  CURRENT FILING DATE:  2003-07-29
;  PRIOR APPLICATION NUMBER: US/09/629,222A
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: 09/463,891
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: PCT/US98/15828
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 60/053,936
;  PRIOR FILING DATE: 1997-07-28
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-629-951-32

  Query Match             22.8%;  Score 697;  DB 15;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-48;
  Matches  126;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        455 DPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAK 514
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAK 60

Qy        515 TIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENH 574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENH 120

Qy        575 EKLSLS 580
              ||||||
Db        121 EKLSLS 126


RESULT 13
US-10-629-951-30
; Sequence 30, Application US/10629951
; Publication No. US20040018550A1
; GENERAL INFORMATION:
;  APPLICANT: Bellacosa, Alfonso
;  TITLE OF INVENTION: Methods for Detection of Transition
;  TITLE OF INVENTION:  Single-Nucleotide Polymorphisns
;  FILE REFERENCE: FCCC 96-21
;  CURRENT APPLICATION NUMBER: US/10/629,951
;  CURRENT FILING DATE:  2003-07-29
;  PRIOR APPLICATION NUMBER: US/09/629,222A
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: 09/463,891
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: PCT/US98/15828
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 60/053,936
;  PRIOR FILING DATE: 1997-07-28
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 119
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-629-951-30

  Query Match             20.2%;  Score 616;  DB 15;  Length 119;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-41;
  Matches  119;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         36 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVK 95
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVK 60

Qy         96 QRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSR 154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSR 119


RESULT 14
US-10-629-951-37
; Sequence 37, Application US/10629951
; Publication No. US20040018550A1
; GENERAL INFORMATION:
;  APPLICANT: Bellacosa, Alfonso
;  TITLE OF INVENTION: Methods for Detection of Transition
;  TITLE OF INVENTION:  Single-Nucleotide Polymorphisns
;  FILE REFERENCE: FCCC 96-21
;  CURRENT APPLICATION NUMBER: US/10/629,951
;  CURRENT FILING DATE:  2003-07-29
;  PRIOR APPLICATION NUMBER: US/09/629,222A
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: 09/463,891
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: PCT/US98/15828
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 60/053,936
;  PRIOR FILING DATE: 1997-07-28
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 37
;   LENGTH: 85
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-629-951-37

  Query Match             14.5%;  Score 443;  DB 15;  Length 85;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-27;
  Matches   85;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         69 IASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYL 128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 IASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYL 60

Qy        129 HKNGETSLKPEDFDFTVLSKRGIKS 153
              |||||||||||||||||||||||||
Db         61 HKNGETSLKPEDFDFTVLSKRGIKS 85


RESULT 15
US-09-967-869A-7
; Sequence 7, Application US/09967869A
; Publication No. US20030082552A1
; GENERAL INFORMATION:
;  APPLICANT: WOLFFE, Alan P.
;  APPLICANT:  URNOV, Fyodor
;  APPLICANT:  LAI, Albert
;  APPLICANT:  RASCHKE, Eva
;  TITLE OF INVENTION: MODULATION OF GENE EXPRESSION USING LOCALIZATION
;  TITLE OF INVENTION:  DOMAINS
;  FILE REFERENCE: 8325-0019 / S19
;  CURRENT APPLICATION NUMBER: US/09/967,869A
;  CURRENT FILING DATE:  2001-09-28
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 68
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  MBD4
US-09-967-869A-7

  Query Match             11.7%;  Score 357;  DB 10;  Length 68;
  Best Local Similarity   98.5%;  Pred. No. 6.2e-21;
  Matches   67;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         81 ECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPED 140
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db          1 ECRKSVPCGWERVVKQRLFGKTAGRFDVYFYSPQGLKFRSKSSLANYLHKNGETSLKPED 60

Qy        141 FDFTVLSK 148
              ||||||||
Db         61 FDFTVLSK 68


Search completed: August 22, 2005, 10:12:39
Job time : 164 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 22, 2005, 10:02:09 ; Search time 43 Seconds
(without alignments)
1006.894 Million cell updates/sec

Title: US-10-629-951-2
Perfect score: 3055
Sequence: 1 MGTTGLESLSLGDRGAAPTV.....HKLNKYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3055	100.0	580	3	US-09-327-984A-38	Sequence 38, Appl
2	3055	100.0	580	4	US-09-629-222A-2	Sequence 2, Appli
3	3055	100.0	580	4	US-09-657-013-56	Sequence 56, Appl
4	3011	98.6	574	4	US-09-629-222A-24	Sequence 24, Appl
5	1326	43.4	384	4	US-09-629-222A-29	Sequence 29, Appl
6	1230	40.3	257	4	US-09-205-258-416	Sequence 416, App
7	697	22.8	126	4	US-09-629-222A-32	Sequence 32, Appl
8	616	20.2	119	4	US-09-629-222A-30	Sequence 30, Appl
9	443	14.5	85	4	US-09-629-222A-37	Sequence 37, Appl
10	254	8.3	50	4	US-09-205-258-1100	Sequence 1100, Ap
11	220.5	7.2	467	4	US-09-657-013-69	Sequence 69, Appl
12	220.5	7.2	467	4	US-09-657-013-70	Sequence 70, Appl
13	213.5	7.0	486	4	US-09-657-013-74	Sequence 74, Appl
14	209	6.8	476	4	US-09-657-013-75	Sequence 75, Appl
15	208.5	6.8	477	4	US-09-657-013-63	Sequence 63, Appl
16	208.5	6.8	486	4	US-09-657-013-55	Sequence 55, Appl
17	208.5	6.8	486	4	US-09-657-013-59	Sequence 59, Appl
18	208.5	6.8	486	4	US-09-657-013-60	Sequence 60, Appl
19	208.5	6.8	486	4	US-09-657-013-62	Sequence 62, Appl
20	208.5	6.8	486	4	US-09-657-013-65	Sequence 65, Appl
21	208.5	6.8	486	4	US-09-657-013-68	Sequence 68, Appl
22	208.5	6.8	486	4	US-09-657-013-72	Sequence 72, Appl
23	208.5	6.8	486	4	US-09-657-013-73	Sequence 73, Appl
24	208.5	6.8	502	4	US-09-949-016-10209	Sequence 10209, A
25	202.5	6.6	345	4	US-09-657-013-112	Sequence 112, App
26	201	6.6	492	4	US-09-657-013-66	Sequence 66, Appl
27	198.5	6.5	484	4	US-09-657-013-58	Sequence 58, Appl
28	198.5	6.5	484	4	US-09-657-013-61	Sequence 61, Appl
29	198.5	6.5	484	4	US-09-657-013-67	Sequence 67, Appl
30	198.5	6.5	484	4	US-09-657-013-71	Sequence 71, Appl
31	194	6.4	219	4	US-09-513-999C-6132	Sequence 6132, Ap
32	192	6.3	132	4	US-09-629-222A-31	Sequence 31, Appl
33	186	6.1	92	4	US-09-657-013-64	Sequence 64, Appl
34	146.5	4.8	1579	3	US-08-755-587-184	Sequence 184, App
35	145	4.7	3418	2	US-08-639-501-2	Sequence 2, Appli
36	145	4.7	3418	2	US-08-603-753D-4	Sequence 4, Appli
37	145	4.7	3418	3	US-09-044-946-2	Sequence 2, Appli
38	145	4.7	3418	3	US-08-755-587-44	Sequence 44, Appl
39	145	4.7	3418	3	US-09-044-908-2	Sequence 2, Appli
40	145	4.7	3418	3	US-09-099-753-4	Sequence 4, Appli
41	145	4.7	3418	3	US-08-986-106-4	Sequence 4, Appli
42	144	4.7	2329	3	US-08-755-587-16	Sequence 16, Appl
43	136.5	4.5	598	4	US-09-538-092-1083	Sequence 1083, Ap
44	131.5	4.3	291	4	US-09-657-013-48	Sequence 48, Appl
45	131.5	4.3	291	4	US-09-949-016-6438	Sequence 6438, Ap

ALIGNMENTS

RESULT 1
US-09-327-984A-38
; Sequence 38, Application US/09327984A
; Patent No. 6368594
; GENERAL INFORMATION:
; APPLICANT: Doetsch, Paul W.
; APPLICANT: Kaur, Balveen
; APPLICANT: Avery, Angela M.
; TITLE OF INVENTION: Broad Specificity DNA Damage Endonuclease
; FILE REFERENCE: 25-98
; CURRENT APPLICATION NUMBER: US/09/327,984A
; CURRENT FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: US 60/088,521
; PRIOR FILING DATE: 1998-06-08
; PRIOR APPLICATION NUMBER: US 60/134,752
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-327-984A-38

Query Match 100.0%; Score 3055; DB 3; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.2e-292;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              |||||||||||||||||||||||||||||||||||||||||
Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580


RESULT 2
US-09-629-222A-2
; Sequence 2, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
;  APPLICANT: Bellacosa, Alfonso
;  TITLE OF INVENTION: Methods for Detection of Transition
;  TITLE OF INVENTION:  Single-Nucleotide Polymorphisns
;  FILE REFERENCE: FCCC 96-21
;  CURRENT APPLICATION NUMBER: US/09/629,222A
;  CURRENT FILING DATE:  2000-07-31
;  PRIOR APPLICATION NUMBER: 09/463,891
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: PCT/US98/15828
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 60/053,936
;  PRIOR FILING DATE: 1997-07-28
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 580
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-629-222A-2

  Query Match            100.0%;  Score 3055;  DB 4;  Length 580;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-292;
  Matches  580;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
              |||||||||||||||||||||||||||||||||||||||||
Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580


RESULT 3
US-09-657-013-56
; Sequence 56, Application US/09657013
; Patent No. 6709817
; GENERAL INFORMATION:
;  APPLICANT: Zoghbi, Huda Y.
;  APPLICANT:  Van den Veyver, Ignatia B
;  APPLICANT:  Amir, Ruthie
;  APPLICANT:  Francke, Uta
;  TITLE OF INVENTION: Methods of Identifying Mutations in a Methyl-CPG-Binding Domain
;  TITLE OF INVENTION:  Containing Gene or Protein in Neurodevelopmental Disease and Trea
;  FILE REFERENCE: HO-P01893US1/09905371
;  CURRENT APPLICATION NUMBER: US/09/657,013
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: US 60/152,778
;  PRIOR FILING DATE: 1999-09-07
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
;   LENGTH: 580
;   TYPE: PRT
;   ORGANISM: Human
US-09-657-013-56

  Query Match            100.0%;  Score 3055;  DB 4;  Length 580;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-292;
  Matches  580;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 RSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480
```

```
Db        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Db        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580

Db        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580
```

RESULT 4
US-09-629-222A-24
; Sequence 24, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; TITLE OF INVENTION: Single-Nucleotide Polymorphisns
; FILE REFERENCE: FCCC 96-21
; CURRENT APPLICATION NUMBER: US/09/629,222A
; CURRENT FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-629-222A-24

Query Match 98.6%; Score 3011; DB 4; Length 574;
Best Local Similarity 99.0%; Pred. No. 4.7e-288;
Matches 574; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

```
Qy          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Db          1 MGTTGLESLSLGDRGAAPTVTSSERLVPDPPNDLRKEDVAMELERVGEDEEQMMIKRSSE 60

Qy         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Db         61 CNPLLQEPIASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRS 120

Qy        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Db        121 KSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRT 180

Qy        181 RSKCKKDVFMPPSSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Db        181 RSKCKKDVFMPPSSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKG 240

Qy        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Db        241 IPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Db        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360

Qy        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSS 420

Db        361 KVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT------EDTIPRTQIERRKTSLYFSS 414

Qy        421 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 480

Db        415 KYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWK 474

Qy        481 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 540

Db        475 FLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELHGIGKY 534

Qy        541 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 580

Db        535 GNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENHEKLSLS 574
```

RESULT 5
US-09-629-222A-29
; Sequence 29, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
; APPLICANT: Bellacosa, Alfonso
; TITLE OF INVENTION: Methods for Detection of Transition
; TITLE OF INVENTION: Single-Nucleotide Polymorphisns
; FILE REFERENCE: FCCC 96-21
; CURRENT APPLICATION NUMBER: US/09/629,222A
; CURRENT FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/463,891
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US98/15828
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 60/053,936
; PRIOR FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-629-222A-29

Query Match 43.4%; Score 1326; DB 4; Length 384;
Best Local Similarity 54.3%; Pred. No. 5.3e-122;
Matches 283; Conservative 29; Mismatches 71; Indels 138; Gaps 4;

```
Qy         36 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVK 95

Db          1 KEDIAVGLGGVGEDGKDLVI--SSERSSLLQEPTAST-LSSTTATEGHKPVPCGWERVVK 57

Qy         96 QRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRY 155

Db         58 QRLSGKTAGKFDVYFISPQGLKFRSKRSLANYLLKNGETFLKPEDFNFTVLPKGSINPGY 117

Qy        156 KDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSSELQESRGLSNFTSTHLLLK 215

Db        118 KHQSLAALTSLQPNETDVSKQNLKTRSKWKTDVLPLPSGTSESPESSGLSNSNSACLLLR 177

Qy        216 EDEGVDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAES 275

Db        178 EHRDIQDVDSEKRRKSKRKVTVLKGTASQKTKQKCRKSLLESTQRNRKRAS--------- 228

Qy        276 EPVAQKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINK 335

Db        229 ------------------------------------------------------------ 228

Qy        336 FCSAKDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTG 395

Db        229 ------------------------------------------------------------ 228

Qy        396 EKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAFKKWTPPRSPFNLVQETLFHD 455

Db        229 -----EDSIPRTQVEKRKTSLYFSSKYNKEALSPPRRKSFKKWTPPRSPFNLVQEILFHD 283

Qy        456 PWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAKT 515

Db        284 PWKLLIATIFLNRTSGKMAIPVLWEFLEKYPSAEVARAADWRDVSELLKPLGLYDLRAKT 343

Qy        516 IVKFSDEYLTKQWKYPIELHGIG-KYGNDSYRIFCVNEWKQ 555

Db        344 IIKFSDEYLTKQWRYPIELHGIWLKYGNDSYRIFCVNEWKQ 384
```

```
RESULT 6
US-09-205-258-416
; Sequence 416, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
;  APPLICANT: Young et al.
;  TITLE OF INVENTION: 207 Human Secreted Proteins
;  FILE REFERENCE: PZ007P1
;  CURRENT APPLICATION NUMBER: US/09/205,258
;  CURRENT FILING DATE: 1998-12-04
;  EARLIER APPLICATION NUMBER: PCT/US98/11422
;  EARLIER FILING DATE: 1998-06-04
;  EARLIER APPLICATION NUMBER: 60/048,885
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,375
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,881
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,880
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,896
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,020
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,876
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,895
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,884
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,894
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,971
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,882
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,899
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,893
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,900
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,901
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,892
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,915
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,019
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,970
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,972
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,916
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,373
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,875
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,374
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,917
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,949
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,974
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,883
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,897
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,898
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,962
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,963
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,877
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,878
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/070,923
;  EARLIER FILING DATE: 1997-12-18
;  EARLIER APPLICATION NUMBER: 60/092,921
;  EARLIER FILING DATE: 1998-07-15
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  NUMBER OF SEQ ID NOS: 1227
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 416
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (100)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-205-258-416

  Query Match             40.3%;  Score 1230;  DB 4;  Length 257;
  Best Local Similarity   93.1%;  Pred. No. 8.2e-113;
  Matches  241;  Conservative    3;  Mismatches    9;  Indels    6;  Gaps    1;

Qy        160 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAALTSHLQNQSNNSNWNLRTRSKCKKDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEG 60

Qy        220 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVA 279
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 VDDVNFRKVRKPKGKVTILKGIPIKKTKKGCRKSCSGFVXSDSKRESVCNKADAESEPVA 120

Qy        280 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QKSQLDRTVCISDAGACGETLSVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSA 180

Qy        340 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFTGEKIF 399
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KDSEHNEKYEDTFLESEEIGTKVEVVERKEHLHTDILKRGSEMDNNCSPTRKDFT----- 235

Qy        400 QEDTIPRTQIERRKTSLYF 418
              ||||||    :::    |
Db        236 -EDTIPRNTDRKKENKPVF 253


RESULT 7
US-09-629-222A-32
; Sequence 32, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
;  APPLICANT: Bellacosa, Alfonso
;  TITLE OF INVENTION: Methods for Detection of Transition
;  TITLE OF INVENTION:  Single-Nucleotide Polymorphisns
;  FILE REFERENCE: FCCC 96-21
;  CURRENT APPLICATION NUMBER: US/09/629,222A
;  CURRENT FILING DATE:  2000-07-31
;  PRIOR APPLICATION NUMBER: 09/463,891
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: PCT/US98/15828
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 60/053,936
;  PRIOR FILING DATE: 1997-07-28
```

```
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-629-222A-32

  Query Match             22.8%;  Score 697;  DB 4;  Length 126;
  Best Local Similarity   100.0%;  Pred. No. 1e-60;
  Matches  126;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        455 DPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAK 514
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 DPWKLLIATIFLNRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSELLKPLGLYDLRAK 60

Qy        515 TIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENH 574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TIVKFSDEYLTKQWKYPIELHGIGKYGNDSYRIFCVNEWKQVHPEDHKLNKYHDWLWENH 120

Qy        575 EKLSLS 580
              ||||||
Db        121 EKLSLS 126


RESULT 8
US-09-629-222A-30
; Sequence 30, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
;  APPLICANT: Bellacosa, Alfonso
;  TITLE OF INVENTION: Methods for Detection of Transition
;  TITLE OF INVENTION:  Single-Nucleotide Polymorphisns
;  FILE REFERENCE: FCCC 96-21
;  CURRENT APPLICATION NUMBER: US/09/629,222A
;  CURRENT FILING DATE:  2000-07-31
;  PRIOR APPLICATION NUMBER: 09/463,891
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: PCT/US98/15828
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 60/053,936
;  PRIOR FILING DATE: 1997-07-28
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 119
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-629-222A-30

  Query Match             20.2%;  Score 616;  DB 4;  Length 119;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-53;
  Matches  119;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         36 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVK 95
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKSVPCGWERVVK 60

Qy         96 QRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSR 154
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 QRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSR 119


RESULT 9
US-09-629-222A-37
; Sequence 37, Application US/09629222A
; Patent No. 6599700
; GENERAL INFORMATION:
;  APPLICANT: Bellacosa, Alfonso
;  TITLE OF INVENTION: Methods for Detection of Transition
;  TITLE OF INVENTION:  Single-Nucleotide Polymorphisns
;  FILE REFERENCE: FCCC 96-21
;  CURRENT APPLICATION NUMBER: US/09/629,222A
;  CURRENT FILING DATE:  2000-07-31
;  PRIOR APPLICATION NUMBER: 09/463,891
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: PCT/US98/15828
;  PRIOR FILING DATE: 1998-07-28
;  PRIOR APPLICATION NUMBER: 60/053,936
;  PRIOR FILING DATE: 1997-07-28
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 37
;   LENGTH: 85
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-629-222A-37

  Query Match             14.5%;  Score 443;  DB 4;  Length 85;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-36;
  Matches   85;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         69 IASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYL 128
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 IASAQFGATAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYL 60

Qy        129 HKNGETSLKPEDFDFTVLSKRGIKS 153
              |||||||||||||||||||||||||
Db         61 HKNGETSLKPEDFDFTVLSKRGIKS 85


RESULT 10
US-09-205-258-1100
; Sequence 1100, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
;  APPLICANT: Young et al.
;  TITLE OF INVENTION: 207 Human Secreted Proteins
;  FILE REFERENCE: PZ007P1
;  CURRENT APPLICATION NUMBER: US/09/205,258
;  CURRENT FILING DATE: 1998-12-04
;  EARLIER APPLICATION NUMBER: PCT/US98/11422
;  EARLIER FILING DATE: 1998-06-04
;  EARLIER APPLICATION NUMBER: 60/048,885
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,375
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,881
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,880
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,896
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,020
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,876
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,895
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,884
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,894
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,971
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,964
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,882
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,899
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,893
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,900
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,901
```

```
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,892
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,915
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,019
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,970
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,972
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,916
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,373
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,875
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/049,374
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,917
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,949
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,974
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,883
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,897
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,898
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,962
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,963
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,877
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/048,878
;  EARLIER FILING DATE: 1997-06-06
;  EARLIER APPLICATION NUMBER: 60/070,923
;  EARLIER FILING DATE: 1997-12-18
;  EARLIER APPLICATION NUMBER: 60/092,921
;  EARLIER FILING DATE: 1998-07-15
;  EARLIER APPLICATION NUMBER: 60/094,657
;  EARLIER FILING DATE: 1998-07-30
;  NUMBER OF SEQ ID NOS: 1227
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1100
;   LENGTH: 50
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-205-258-1100

  Query Match             8.3%;  Score 254;  DB 4;  Length 50;
  Best Local Similarity   98.0%;  Pred. No. 1.2e-17;
  Matches   49;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        110 FISPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCS 159
              | |||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FSSPQGLKFRSKSSLANYLHKNGETSLKPEDFDFTVLSKRGIKSRYKDCS 50


RESULT 11
US-09-657-013-69
; Sequence 69, Application US/09657013
; Patent No. 6709817
; GENERAL INFORMATION:
;  APPLICANT: Zoghbi, Huda Y.
;  APPLICANT:  Van den Veyver, Ignatia B
;  APPLICANT:  Amir, Ruthie
;  APPLICANT:  Francke, Uta
;  TITLE OF INVENTION: Methods of Identifying Mutations in a Methyl-CPG-Binding Domain
;  TITLE OF INVENTION:  Containing Gene or Protein in Neurodevelopmental Disease and Tre
;  FILE REFERENCE: HO-P01893US1/09905371
;  CURRENT APPLICATION NUMBER: US/09/657,013
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: US 60/152,778
;  PRIOR FILING DATE: 1999-09-07
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69
;   LENGTH: 467
;   TYPE: PRT
;   ORGANISM: Frog
US-09-657-013-69

  Query Match             7.2%;  Score 220.5;  DB 4;  Length 467;
  Best Local Similarity   25.7%;  Pred. No. 1e-12;
  Matches   98;  Conservative   60;  Mismatches  157;  Indels   67;  Gaps   15;

Qy         22 SSERLVPDPPNDLRKEDVAMELER---VGEDEEQMMIKRS--SECNPLLQEPIASAQFGA 76
              |||   |  | | | |::   |    | |       :||   :  |: ::|
Db         49 SSEH-QPGEPADEGKADMSESAEENLAVPESSASPKQRRSVIRDRGPMYEDP-------- 99

Qy         77 TAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSL 136
                 ::| || | :||  |::||:|||| |:| |  |||| |  |  | |:|||
Db        100 --------TLPEGWTRKLKQRKSGRSAGKFDVYLINPNGKAFRSKVELIAYFQKVGDTSL 151

Qy        137 KPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDV--FMPPSS 194
               | |||||| : ||  || :           : ::  |: : | | : | :    ||
Db        152 DPNDFDFTV-TGRGSPSRREQ------KQPKKPKAPKSSVSGRGRGRPKGSIKKVKPPVK 204

Qy        195 SSELQESRGLSNFTSTHLLLK-----EDEGVDDVNFRKV-------RKPKGKVTILKGIP 242
              |  :|  | :   :   ||:|       |  |    ::|       || | : |    |
Db        205 SEGVQVKRVIEK-SPGKLLVKMPYSGTKEASDATTSQQVLVIKRGGRKRKSE-TDPSAAP 262

Qy        243 IKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLSV 302
               |: :|    | :      :|::::    ::  :|: :      ||         ||:||
Db        263 KKRGRKPSNVSLAAAAAEAAKKKAI---KESSIKPLLE------TVLPIKKRKTRETISV 313

Qy        303 TSEE-------NSLVKK------KERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYE 349
                ::         :::|        :|  | |   | :  :|:  |        |:  :
Db        314 DVKDTIKPEPLTPVIEKVMKGQNPAKSPESRSTEGSPKIKTGLPKKELQQHHHHHHHHH 373

Qy        350 DTFLESEEIGTKVEVVERKEHL 371
                ||:   |  |    |:::
Db        374 HHHSESKASATSPEPETSKDNI 395


RESULT 12
US-09-657-013-70
; Sequence 70, Application US/09657013
; Patent No. 6709817
; GENERAL INFORMATION:
;  APPLICANT: Zoghbi, Huda Y.
;  APPLICANT:  Van den Veyver, Ignatia B
;  APPLICANT:  Amir, Ruthie
;  APPLICANT:  Francke, Uta
;  TITLE OF INVENTION: Methods of Identifying Mutations in a Methyl-CPG-Binding Domain
;  TITLE OF INVENTION:  Containing Gene or Protein in Neurodevelopmental Disease and Trea
;  FILE REFERENCE: HO-P01893US1/09905371
;  CURRENT APPLICATION NUMBER: US/09/657,013
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: US 60/152,778
;  PRIOR FILING DATE: 1999-09-07
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
;   LENGTH: 467
;   TYPE: PRT
;   ORGANISM: Frog
US-09-657-013-70

  Query Match             7.2%;  Score 220.5;  DB 4;  Length 467;
  Best Local Similarity   25.7%;  Pred. No. 1e-12;
  Matches   98;  Conservative   60;  Mismatches  157;  Indels   67;  Gaps   15;
```

```
Qy          22 SSERLVPDPPNDLRKEDVAMELER---VGEDEEQMMIKRS--SECNPLLQEPIASAQFGA 76
Db          49 SSEH-QPGEPADEGKADMSESAEENLAVPESSASPKQRRSVIRDRGPMYEDP-------- 99

Qy          77 TAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETSL 136
Db         100 --------TLPEGWTRKLKQRKSGRSAGKFDVYLINPNGKAFRSKVELIAYFQKVGDTSL 151

Qy         137 KPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDV--FMPPSS 194
Db         152 DPNDFDFTV-TGRGSPSRREQ------KQPKKPKAPKSSVSGRGRGRPKGSIKKVKPPVK 204

Qy         195 SSELQESRGLSNFTSTHLLLK-----EDEGVDDVNFRKV-------RKPKGKVTILKGIP 242
Db         205 SEGVQVKRVIEK-SPGKLLVKMPYSGTKEASDATTSQQVLVIKRGGRKRKSE-TDPSAAP 262

Qy         243 IKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLSV 302
Db         263 KKRGRKPSNVSLAAAAAAEAAKKKAI---KESSIKPLLE------TVLPIKKRKTRETISV 313

Qy         303 TSEE-------NSLVKK------KERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYE 349
Db         314 DVKDTIKPEPLTPVIEKVMKGQNPAKSPESRSTEGSPKIKTGLPKKELQQHHHHHHHHH 373

Qy         350 DTFLESEEIGTKVEVVERKEHL 371
Db         374 HHHSESKASATSPEPETSKDNI 395
```

RESULT 13
US-09-657-013-74
; Sequence 74, Application US/09657013
; Patent No. 6709817
; GENERAL INFORMATION:
; APPLICANT: Zoghbi, Huda Y.
; APPLICANT: Van den Veyver, Ignatia B
; APPLICANT: Amir, Ruthie
; APPLICANT: Francke, Uta
; TITLE OF INVENTION: Methods of Identifying Mutations in a Methyl-CPG-Binding Domain
; TITLE OF INVENTION: Containing Gene or Protein in Neurodevelopmental Disease and Tre
; FILE REFERENCE: HO-P01893US1/09905371
; CURRENT APPLICATION NUMBER: US/09/657,013
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 60/152,778
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Human
US-09-657-013-74

Query Match 7.0%; Score 213.5; DB 4; Length 486;
Best Local Similarity 24.4%; Pred. No. 5.3e-12;
Matches 106; Conservative 53; Mismatches 147; Indels 129; Gaps 18;

```
Qy          37 EDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG----------------ATAGT 80
Db          22 KDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASP 81

Qy          81 ECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANY 127
Db          82 KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAY 141

Qy         128 LHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKD 187
Db         142 FEKVGDTSLDPNDFDFTV-TGRGSPSR-----------------------REQKPPKK- 175

Qy         188 VFMPPSSSSELQESRGL---SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIK 244
Db         176 ----PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF- 226

Qy         245 KTKKGCRKSCSGFVQSDS----KRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
Db         227 QTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKR----------GRKPGSVV 276

Qy         301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
Db         277 AAAAAE---AKKK---AVKGSSIRSVQETVLPIKK----------RKTRET--------V 312

Qy         361 KVEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYF 418
Db         313 SIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSPKGR---------- 354

Qy         419 SSKYNKEALSPPRRK 433
Db         355 ----SSSASSPPKKE 365
```

RESULT 14
US-09-657-013-75
; Sequence 75, Application US/09657013
; Patent No. 6709817
; GENERAL INFORMATION:
; APPLICANT: Zoghbi, Huda Y.
; APPLICANT: Van den Veyver, Ignatia B
; APPLICANT: Amir, Ruthie
; APPLICANT: Francke, Uta
; TITLE OF INVENTION: Methods of Identifying Mutations in a Methyl-CPG-Binding Domain
; TITLE OF INVENTION: Containing Gene or Protein in Neurodevelopmental Disease and Tre
; FILE REFERENCE: HO-P01893US1/09905371
; CURRENT APPLICATION NUMBER: US/09/657,013
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 60/152,778
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Human
US-09-657-013-75

Query Match 6.8%; Score 209; DB 4; Length 476;
Best Local Similarity 24.0%; Pred. No. 1.4e-11;
Matches 104; Conservative 53; Mismatches 149; Indels 128; Gaps 17;

```
Qy          37 EDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG----------------ATAGTE 81
Db          13 KDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSARLCEASASPK 72

Qy          82 CRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYL 128
Db          73 QRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYF 132

Qy         129 HKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKDV 188
Db         133 EKVGDTSLDPNDFDFTV-TGRGSPSR-----------------------REQKPPKK-- 165

Qy         189 FMPPSSSSELQESRGL---SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIKK 245
Db         166 ---PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF-Q 217

Qy         246 TKKGCRKSCSGFVQSDS----KRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETLS 301
Db         218 TSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKR----------GRKPGSVVA 267

Qy         302 VTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGTK 361
Db         268 AAAAEAKKKAVKESSIR------SVQETVLPIKK----------RKTRET--------VS 303

Qy         362 VEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFS 419
Db         304 IEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSPKGR---------- 344
```

```
Qy        420 SKYNKEALSPPRRK 433
                  :  | |||:::
Db        345 ---SSSASSPPKKE 355


RESULT 15
US-09-657-013-63
; Sequence 63, Application US/09657013
; Patent No. 6709817
; GENERAL INFORMATION:
;  APPLICANT: Zoghbi, Huda Y.
;  APPLICANT:  Van den Veyver, Ignatia B
;  APPLICANT:  Amir, Ruthie
;  APPLICANT:  Francke, Uta
;  TITLE OF INVENTION: Methods of Identifying Mutations in a Methyl-CPG-Binding Domain
;  TITLE OF INVENTION:  Containing Gene or Protein in Neurodevelopmental Disease and Tre
;  FILE REFERENCE: HO-P01893US1/09905371
;  CURRENT APPLICATION NUMBER: US/09/657,013
;  CURRENT FILING DATE:  2000-09-07
;  PRIOR APPLICATION NUMBER: US 60/152,778
;  PRIOR FILING DATE: 1999-09-07
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63
;   LENGTH: 477
;   TYPE: PRT
;   ORGANISM: Human
US-09-657-013-63

  Query Match             6.8%;  Score 208.5;  DB 4;  Length 477;
  Best Local Similarity   23.9%;  Pred. No. 1.6e-11;
  Matches  104;  Conservative   53;  Mismatches  149;  Indels  129;  Gaps   17;

Qy         37 EDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFG----------------ATAGT 80
              :|  :: ::| :|:::    :     |        |: |                |:|
Db         13 KDKPLKFKKVKKDKKEEKEGKHEPVQPSAHHSAEPAEAGKAETSEGSGSAPAVPEASASP 72

Qy         81 ECRKSV-------------PCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANY 127
              : |:|:             | || | :|||  |::||::||| |:|||  ||||  |  |
Db         73 KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAY 132

Qy        128 LHKNGETSLKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCKKD 187
                | |:||| | |||||| : ||  ||                          | :   ||
Db        133 FEKVGDTSLDPNDFDFTV-TGRGSPSR-----------------------REQKPPKK- 166

Qy        188 VFMPPSSSSELQESRGL---SNFTSTHLLLKEDEGVDDVNFRKVRKPKGKVTILKGIPIK 244
                | |        ||        :|        |||     | : |  ||  :|  :|
Db        167 ----PKSPKAPGTGRGRGRPKGSGTTRPKAATSEGVQVK--RVLEKSPGK--LLVKMPF- 217

Qy        245 KTKKGCRKSCSGFVQSDS----KRESVCNKADAESEPVAQKSQLDRTVCISDAGACGETL 300
              :|  | :    |   |      ||     ||:|: : : :|           |  :
Db        218 QTSPGGKAEGGGATTSTQVMVIKRPGRKRKAEADPQAIPKKR----------GRKPGSVV 267

Qy        301 SVTSEENSLVKKKERSLSSGSNFCSEQKTSGIINKFCSAKDSEHNEKYEDTFLESEEIGT 360
              :  : |       || |:        | |:|   | |           |  :|
Db        268 AAAAAEAKKKAVKESSIR------SVQETVLPIKK----------RKTRET--------V 303

Qy        361 KVEVVERKEHLHTDIL--KRGSEMDNNCSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYF 418
               :|| |  : |    |  | |  :    || ||         ::: |: :
Db        304 SIEVKEVVKPLLVSTLGEKSGKGLKTCKSPGRK--------SKESSPKGR---------- 345

Qy        419 SSKYNKEALSPPRRK 433
                  :  | |||:::
Db        346 ----SSSASSPPKKE 356


Search completed: August 22, 2005, 10:09:50
Job time : 44 secs
```